



森林資源調查測計研究室

# 自然資源地理資訊系統學

"土地資源評估的地理資訊系統原則"

(Principles of GIS for Land Resources Assessment)

著 者：馮豐隆  
高堅泰  
鄭吉延  
簡炯欣  
整理者：陳淑芬

中興大學森林學系

1999. 3. 4



# "土地資源評估的地理資訊系統原則" *Principles of GIS for Land Resources Assessment*

## 目 錄

序	馮豐隆	1
第一章	地理資訊系統 馮豐隆 編	4
第二章	主題圖的資料結構 簡炯欣 編	13
第三章	數位高層模型 高堅泰 編	36
第四章	資料的輸入、驗證、儲存與輸出 鄭吉延 編	50
第五章	資料分析及空間模式建立的方法 高堅泰 編	62
第六章	資料品質、誤差及自然變異 簡炯欣 編	75
第七章	分類方法 鄭吉延 編	95
第八章	空間內插的方法 高堅泰、鄭吉延 編	109
第九章	如何選擇 GIS 簡炯欣 編	124
附 錄		130

## 土地資源評估的地理資訊系統原則

Burrough P.A.1993 Principles of GIS for Land Resources Assessment  
Clarendon Press, Oxford

一、本書主要在描述與說明 GIS 的原理，其對於 GIS 應用於環境與自然資源調查與分析是非常重要的。GIS 主要的內容包括：網格與向量式資料結構、資料輸入模組、驗證、貯存和產出、DTM、空間分析方法和模式以及分類方法與插入方法。且對於許多 GIS 共同使用的原則加以解釋。另外也談論 GIS 最近的發展、GIS 在資源經營管理應用的回顧、GIS 處理過程中產生的誤差與延展。雖然本教科書引用了許多應用 GIS 所產生的結果，然本書重點並不是對目前 GIS 應用的回顧，而是要利用這些成果來解釋隱藏在目前常使用系統內的原理。若讀者想學習 GIS 的特殊應用的話，可參考 Teicholz and Berry(1983), Tomlinson *et al.* (1976), Salmen *et al.* (1976), Knapp and Rider (1979), Nagy and Wagle (1979), Marble *et al.* (1981), Dueker (1979), Burrough and de Veer (1984), Fabos and Caswell (1972), Fraser-Taylor (1980)(可參考 preface!)哈佛大學的電腦製圖與空間分析研究室(the Laboratory for Computer Graphics and Spatial Analysis, Harvard University)、國際製圖聯合會(International Cartographical Union)AUTOCARTO 會議論文集、土壤資訊系統工作小組出版品(Moore *et al.* 1981, Burrough and Bie 1984)。

## 二、本書的主要組成包括

### (一)地理資訊系統

- 1.前言—地圖與空間資訊
- 2.電腦輔助製圖與圖籍分析
- 3.地理資訊系統
- 4.GIS 的組成
- 5.GIS 未來的方向與趨勢
- 6.參考文獻

### (二)主題圖的資料結構

- 1.GIS 的資料結構
- 2.點、線、面
- 3.地圖的定義
- 4.電腦的地理資料
- 5.資料庫結構—電腦的資料組織
- 6.檔案與資料的獲得
- 7.資料庫結構
- 8.地理資料的結構與電腦的表現
- 9.網格資料
- 10.地理單元的向量資料結構

11.主題圖的資料結構—網格與向量的選擇

12.參考文獻

(三)數位高程模型 (Digital Elevation Models, DEM)

- 1.對 DEM 的需求
- 2.表現 DEM 的方法
- 3.影像方法
- 4.DEMs 的資料來源與取樣方法
- 5.可由 DEMs 導出的產物
- 6.由 DEMs 自動地形描繪
- 7.參考文獻

(四)資料輸入、驗證、貯存與產出

- 1.資料輸入
- 2.資料的驗證、修正與貯存
- 3.資料產出
- 4.使用者介面
- 5.參考文獻

(五)資料分析的方法與空間模式發展

- 1.前言
- 2.資料庫的定義
- 3.簡單資料的擷取
- 4.疊圖的一般方法
- 5.使用自然語言指令的繪圖模型
- 6.繪圖模式的連結指令
- 7.土地評估、規劃、繪圖模式的優缺點
- 8.參考文獻

(六)資料品質、誤差與自然變異

- 1.誤差的(明顯)來源
- 2.由自然變異或原始量測產出的誤差
- 3.由於處理所產的誤差
- 4.由於疊合和邊界交叉所產生的誤差與問題
- 5.向量圖籍網格化所產生的誤差
- 6.與數化圖籍或地理標註有關的誤差
- 7.疊合二個或多個多邊形所產生的誤差
- 8.邊界的性質
- 9.邊界的統計性質
- 10.疊合圖籍的組合屬性
- 11.結論
- 12.參考文獻

(七)分類的方法

- 1.分類
- 2.多變值分析與分類
- 3.分派單元至已存在的等級裡
- 4.為 GIS 的專家系統
- 5.GIS 的分類方法 — 結論
- 6.參考文獻

(八)空間插入的方法

- 1.插入目前可支使用的方法
- 2.插入的區域方法
- 3.地域插入
- 4.使用空間與變異的適當插入法
- 5.Kriging 將點、線擴展至大面積
- 6.比較 Kniging 與其他插入技術
- 7.參考文獻

(九)選擇地理資訊系統

- 1.定義 GIS 的需求
- 2.設立 GIS 的步驟
- 3.參考文獻

附錄一：使用於 GIS 的專有名詞

附錄二：GIS 的來源一覽表

作者索引

主題索引

三、GIS 的有關文獻分散於許多不同領域的雜誌裡，本書儘可能的將讀書所需的目錄(bibliography)，列出供參考。

電腦硬體價格與作業的效率有相當的關係，目前價格下降不少，但與軟體一樣依然需要使用者一筆很大的投資，目前技術更新很快，但原則方面的討論仍相當有效，軟、硬體如何有效搭配以符合組織的需求，是購置 GIS 最主要的考慮，所以於本書的最後一章節討論。

往後可利用 WWW 學習 (Keyword) GIS。

# 第一章 地理資訊系統(Geographic Information System, GIS)

## (一)簡介：地圖及空間資訊

整理地表空間分布重要性質的資料已成為組織社會一項重要活動。由古文明迄今，空間資料係由航行家、地理學家、調查者收集，交由製圖者(cartographer)繪製。一般都是為了特殊目的而繪製，如：地形圖、疆域圖、航海圖等各種主題圖(thematic maps)，以方便於針對單一主題的瞭解與掌握。十九世紀，歐洲強權擴展其勢力以製圖來評估與瞭解其資源—地質、地形、土壤、生態與土地。

主題圖應用極廣，不僅用於展示一般目的，亦可顯示特殊目的(Fisher 1978, Hodgkiss 1981)。主題可分為定質化、定量化。定質、定量可以利用如圖 1-1 的地誌圖(choropleth map)將同質區歸類，異質區則以邊界區分，如土壤圖、土地利用型圖，普查(censuses)結果的展示圖。定量資料則可用數學來描述連續表面，將等數值的點連接成等值線(isoline or contours)來表示，如圖 1-2 等高線圖與等溫線、等壓線等。

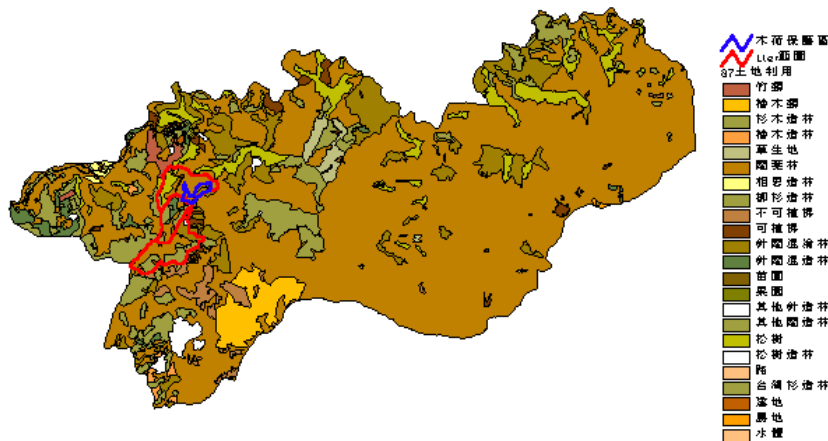


圖 1-1 惠蓀林場的土地利用型圖為地誌圖

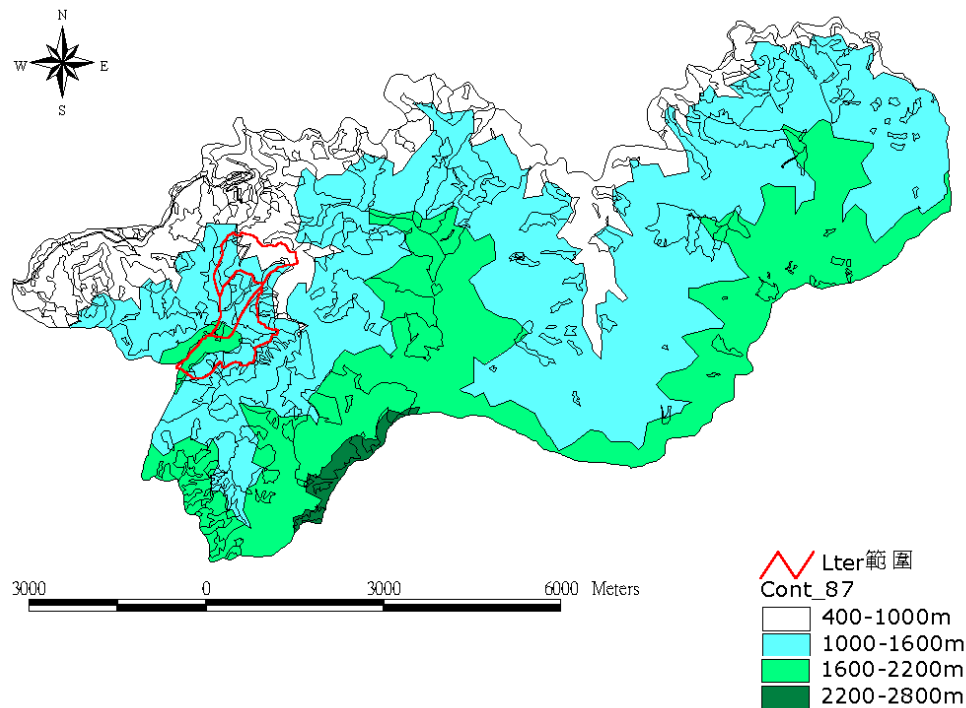


圖 1-2 惠孫林場之等高線圖為等值線圖

20 世紀，對地圖的需求大量增加，立體空中照相及遙感探測使得地圖涵蓋面積更大，也更為正確，這也使得主題圖成為資源開發管理更有用資料的來源。

許多調查的目標就是資源調查，清點存貨(Inventory)，資源調查為量測觀察、分類與記錄的過程，首先發展以適當數學解決空間問題是在 1930-1940 年間，那也正是統計方法、時間序列同時發展的時候。直到 1960 年代，數位電腦利用空間分析的概念方法以及將量化主題圖和空間分析實際加以應用(Cliff and Ord 1981, Journel and Huijbregts 1978, Ripley 1981)。近年來，空間資料及空間分析之需求並不僅地球科學家所重視而已，都市規劃者、地政單位也廣泛應用。

在電腦應用到繪圖以前，所有繪製空間資料地圖皆以點表示，資料是畫在一張紙或投影片上，後來空間資料就以點、線、面表示，以不同符號、顏色、文字代表不同意義，用圖例來說明。

因為紙上地圖伴隨著資料庫的紀要，對資訊收集、編譯及使用，有以下幾個非常重要的結果：

1. 原始資料大量減少、類別簡化，為了使它們易於了解及可敘述，許多地區性的詳細資料都被略過且遺失。
2. 地圖可以畫得很正確，描述得清晰。尤其複雜的主題圖。
3. 表示與地圖尺度有關的面積大小，只能由一些地圖來表示。資訊量少意即與地圖比例尺一致，則僅能由一些圖表示。
4. 資料一旦放入圖中，要與其他空間資料結合並不便宜也不容易。
5. 印出的地圖是靜態、定性的。沒有再收集貯存新資料，而僅以主題圖線劃單元來做定量空間分析是很困難。

收集和編譯資料及出版印製地圖是花錢且耗時的事，同樣地，若由一通用圖中要抽出單一主題的地圖，則要重新用手畫出，當然也是非常昂貴。

當需要地球表面變化的資訊，則傳統的地圖無法表現，然現在可由航空照片，尤其是衛星影像，看出隨時間變化的情形，但其產物非地圖而是航照影像(photographic image)或磁帶上的一連串資料，需要以新工具來把這些數值轉換到地圖上，結合遙測(remote sensing)、地面調查(earth bound survey)及製圖學(cartography)才產生所謂的地理資訊系統(GIS)。

## 2.電腦輔助製圖及圖籍分析(Computer—assisted mapping and map analysis)

1960、1970 年代，地圖資訊應用的新趨勢為：資源評估(resource assessment)、土地評價(land evaluation)及規劃(planning)。

要使計算更完全，有二種方法：

(1)形態描述法(gestalt methods)：找出自然發生的環境單位，將這些可辨認、唯一且相互獨立的屬性畫出圖來，這些相互獨立組合的環境特性，包括地形、地理、土壤、植生和水，即構成 Vink(1981)所提的相關複合體(Correlative Complex)。這個基本觀念亦應用於整合資源調查(integrated resource surveys)，然而，為許多不同目標而使用整合資源調查結果時，則發現主要問題出在所得到的資料太廣了，無法為了某個特別議題，而由某一地景擷取部分所需的資料。

(2)當單一主題有關資源的地圖可利用時，使用者喜歡找方法把這些有用的資料整合而加以綜觀或重新再分類或再產生所需要的資料。當有各種不同單一領域資源地圖(monodisciplinary resource maps)時，使用者傾向於能從整合各圖籍來看，重新分類或更一般化。在美國，規劃者與地景建築者，則認為可以透過疊合(overlap)不同領域之單項主題圖的透明片，將各種資料整合、結合。當中建立此簡單技術最有名的是 Ian McHarg (McHarg 1969)。利用電腦以統計值繪製整合圖製圖的系統(Synagraphic Mapping System)即是 Fish(1963)利用 Edgar M. Horwood 電腦繪圖的觀念。

利用電腦製圖，把統計值印在紙上，以模式分析資料，再將結果以印表機(line printer)秀出，SYMAP(Synagraphic Mapping System)算是第一個組合製圖過程的電腦程式。

因為 SYMAP, GRID, IMGRID, GEOMAP 和其他程式被設計為提供快速且便宜的網格資料，最初結果只能用粗糙的印表機繪圖，許多製圖者拒絕接受這種地圖。直到 1977 年，電腦繪圖的經驗才算完備。

利用電腦繪圖的理由：

- 1.已存在的地圖可以快速印出；
- 2.已存在的地圖可以便宜印出；
- 3.可符合特定使用者之需；
- 4.不需要特殊技術人力下都可生產地圖；
- 5.在相同資料下可允許不同繪圖表現；



- 6.當資料已是數字型態時，可利用製圖及將最近資料繪入圖中；
- 7.可便利分析資料，當需要統計分析及製圖間交互作用時統計分析結果可用圖籍表現；
- 8.減少以印刷地圖作為資料儲存方式，亦可減少資料分類的影響及資料品質的概述；
- 9.量化執行可做到手繪地圖做不到的事，如三維地圖或立體圖；
- 10.可以清晰說明由選擇和一般化過程完成的不同主題的地圖；
- 11.自動化的引進可做為整個製圖過程之檢討，亦可節省及改善製圖過程。

1970 年代後期，才有大量投資在電腦輔助繪圖軟、硬體的建立及應用上，尤其是北美的政府及個別企業，歐洲的規模比北美小，其中包括瑞典、挪威、丹麥、法國、荷蘭、英國、西德等國。

電腦輔助繪圖並未如預期般直接降低成本，這種新技術的獲得及建立經常是很昂貴的，並且需要受過訓練的員工來作業，且電腦輔助設計的工廠及電腦輔助系統都不一樣，以致投資在建立軟、硬體設備之成本不易回收，成果不易看出，同時昂貴系統的購買者被迫需要僱用程式操作員以適應他們所需的特殊系統，在許多繪圖組織中具有許多不確定性，他們只是更快速的完成現有工作或是提供使用者更多的產品而已。

電腦硬體發展快速，組織內容將自動化應用在安全的地方一校正圖樣、掃描輪廓，準備印製的底片。過去許多繪製者並不了解數值化的圖形資料對於空間問題的分析上是強而有力的。

其實，電腦繪圖在技術建立上是極費力的。過去，數以百計的電腦程式，多數都建立在改進手動方法為自動化，而不是在探究空間資料的掌握。

1960—1970 在電腦應用上有兩個趨勢：(1)既有工作的自動化，強調製圖的正確性及視覺上的品質；(2)強調空間分析，但是犧牲了好的繪圖結果。

### **3.地理資訊系統(Geographic Information System, GIS)**

利用電腦繪製地圖及空間分析在自動擷取資料、資料分析及成果顯示在相關領域。這些領域包括地籍及地形圖等主題圖繪製，土木工程、地質學、數學等空間變異的研究；土壤學、調查與航照技術、鄉村規劃、都市計劃、設施網路設計、遙測、影像分析等歷史的發展都是互相平行的、同時的(如圖 1-3)，軍事上的應用則以套疊技術為主且領銜許多領域的整合應用。

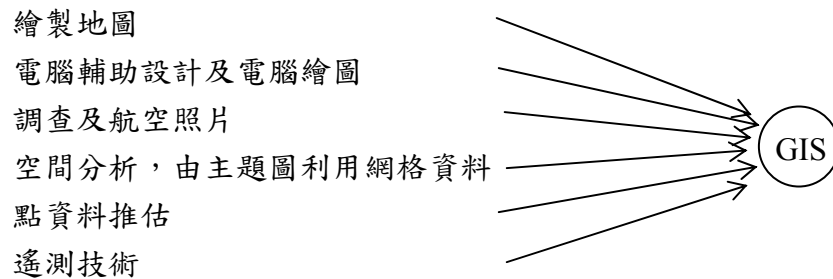


圖 1-3 地理資訊系統是整合不同空間資料處理平行發展的結果

建立一套強有力的系統來收集、儲存、擷取、轉換、展示空間資料就是 GIS。

地理資料建立之條件有三：(1)已知座標系統位置(2)與位置無關的屬性資料，如顏色、PH 價值、罹病率(3)彼此在空間分布位置之關聯。

GIS 與電腦繪圖不同，因後者著重於展示及操縱可看見的物質，而前者則著重空間資料的建立、貯存、分析與展示。現代的 GIS 需要好的電腦繪圖，但一製圖套裝軟體若沒有空間資料的貯存、分析與展示功能，也不能夠獨立有效率地完成工作，繪圖軟體(drawing packages)也無法單獨建立這種系統。

GIS 有許多部分與電腦輔助設計(Computer-Aided Design, CAD)系統所用的資料如由天上的飛機到地面調查等廣泛的範圍，技術的主題都相同，GIS 與 CAD 都需要有目標做參考，都需要非圖形資料，也都需要描述與地形的關係。二者主要的不同在於 GIS 需要大量容量及細分的資料輸入，並使用特別的分析方法，這些不同往往不小，對 CAD 有效的系統也許對 GIS 非常不適合，反之亦然。

#### 4.GIS 的組成(GIS Components)

GIS 的一般硬體組成可由圖 1-4 表示，電腦中央處理系統與磁碟機、數化儀與其他設備則將地圖、文件轉成數量化型式，再送到電腦、繪圖機、印表機，再將成果印製出來或由視覺展示單元(visual display unit, VDU)來監控。

(1)電腦硬體(Computer hardware)：應用軟體單位組、適當組織關聯，三者需達平衡

電腦硬體：

(把資料由地圖數化  
掃描並傳送至電腦)

(展示結果)

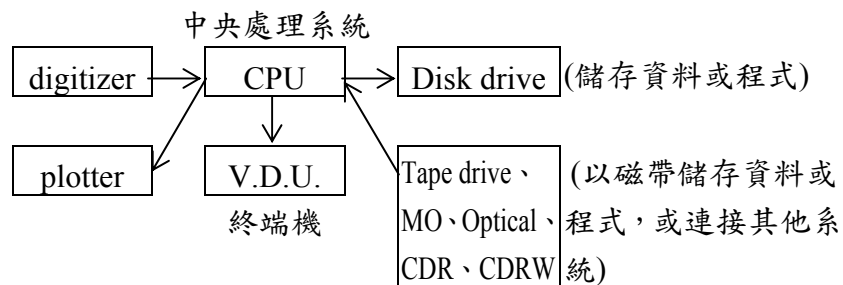


圖 1-4 GIS 主要硬體組成

(2)軟體模組：包括 5 個單位(如圖 1-5)

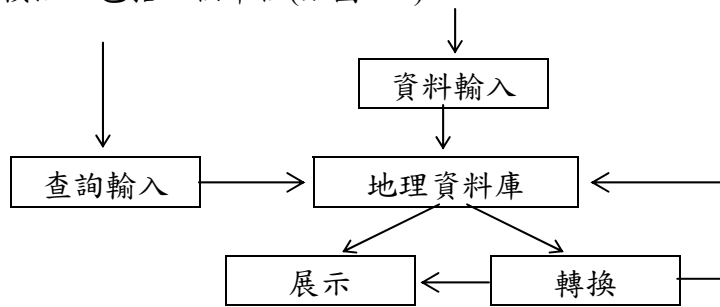


圖 1-5 GIS 主要軟體組成

a. 資料輸入與驗證(data input and verification)：

來源：地圖、野外觀測、感映器(sensor)：包括以航空照相、衛星、紀錄器等收集資料，將資料轉換成數字形態再由終端機、數化儀(digitizer)、掃描器(scanner)以檔案形式貯存於磁片、磁碟內

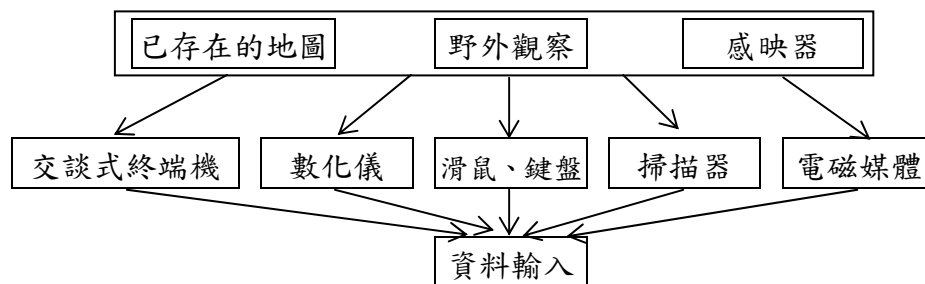


圖 1-6 資料輸入

b. 資料儲存及資料庫管理(Data Storage and Database Management)：

如圖 1-7，資料儲存及資料庫管理系統是將有關位置、位向關係的關連與地理之素的屬性以點、線、面表示在地球表面的物體，予以結構化與組織起來的方式，其必須由電腦來操作且由此系統的使用知悉。

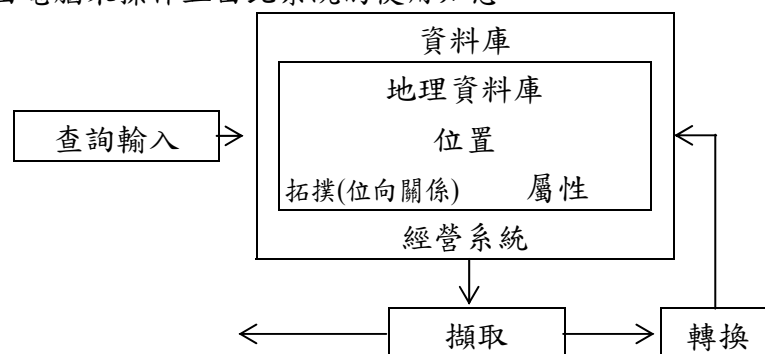


圖 1-7 地理資料庫的組成

c. 資料輸出和描述(Data Output and Presentation)：

資料輸出與描述(如圖 1-8)，係有關資料展示與分析結果的報告的方法。資

料可以地圖、表格和圖形(Figures-Graphs and Charts)等不同方式表示，且透過終端機、印表機、繪圖機，將其資料資料、資訊輸出。

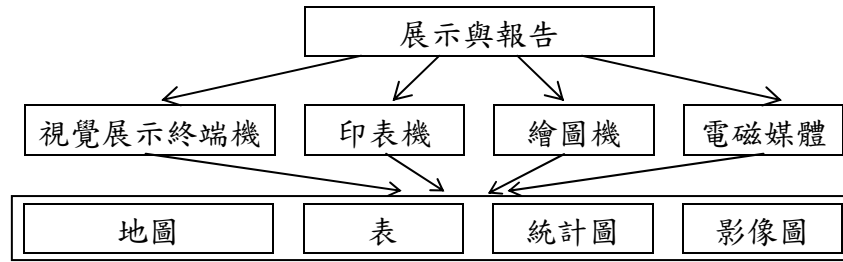


圖 1-8 資料輸出

d. 資料轉換：如圖 1-9 所示包括使用與分析二種操作，也就是(1)轉換需要移除資料中錯誤的部份，需要進行資料更新、需要配合其他要整合應用的資料格式。(2)整序列的分析方法應用到資料上，以達到 GIS 能解答的問題上。

資料轉換觀念可運作在空間及非空間的資料上，不管個別的使用或混合使用都可以。許多資料轉換與尺度(scale)的改變，把資料配合至新的投影座標上，資料的擷取和面積、周界的計算。簡單的轉換方法可以結合以獲得某些地理或空間模型的特定型態是本書主題，在 5,6,7,8 章內涵蓋。有關空間轉換及轉換時確保資料庫的完整，在第 4 章內再加以討論。

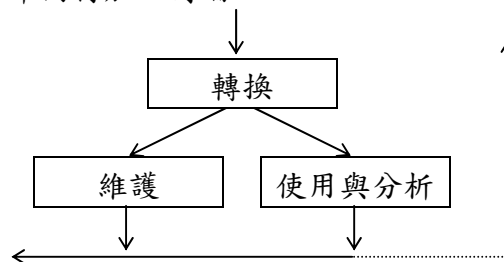


圖 1-9 資料轉換

e. 與使用者的交互關係，GIS 的設計者要解答使用者想問的下列問題：

- (a) A 在那裏？
- (b) 位置 A 與位置 B 有關係的部份在那裏？
- (c) 在 B 的 D 距離內 A 型發生多少？
- (d) 在 X 位置、函數 Z 的值＝？
- (e) B 有多大？(面積、周長、數量)
- (f) 擷取不同類的空間資料結果如何？
- (g) 沿 path P 由 X 到 Y 最小成本、阻力或距離的路徑是？
- (h) 在點  $X_1, X_2, \dots$  是什麼？
- (i) 與某些特性、屬性聯合的下一物件(object)是什麼？
- (j) 重新分類的物件，是由那些特定屬性的組合？
- (k) 在已知假設情境 S 下，在時間 T 中，利用真實世界的數值資料庫做為模型來模擬過程 P？

由傳統方法要回答這些問題是困難的，有些用電腦輔助方法也是困難且耗費

時間。使用者在使用 GIS 時，不智地製造、傳播誤差是很危險，將使工作成為沒價值。這些都在 chap6 有介紹，標準化的 GIS 方法可能導致很大的誤差，需用其他分析方法來降低，這些方法在 chap8 有討論。

與使用者的交互關係—疑問輸入—接受及使用資料系統是絕對必須。這個問題近幾年來才稍為人重視。一般使用者才能直接接觸電腦也是最近的事，由於電腦科技的發達，使電腦走入一般人的生活中。本書假設所有使用者與 GIS 間之關係可經由菜單式指令系統及如英文般的指令而進行。

### 3. GIS 組成面面觀(The organizational aspects of GIS)

GIS 有五個技術的支系統，使其能運作，但其不能保證，所有特定的 GIS 可以有有效的應用，為了使使用有效率，GIS 需置於一適當的組織結構中(如圖 1-10)。

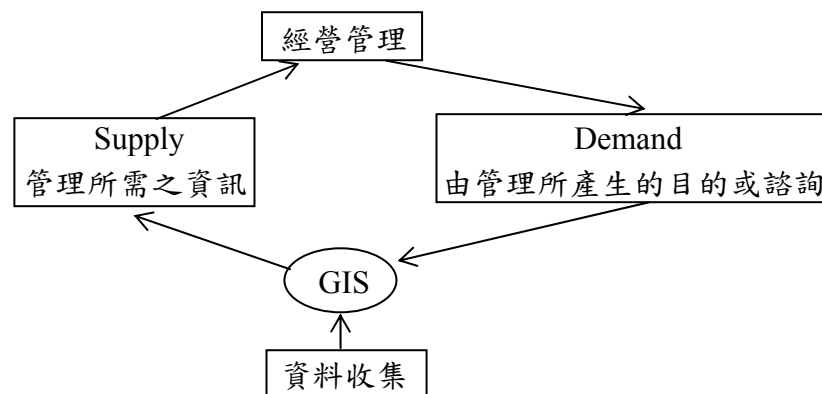


圖 1-10 GIS 的組織

就如同處理複雜產物的組織一樣，新技術唯有完全融入工作流程中而不是事後放馬後炮，才能有效的應用。因此除了投資在軟硬體上之外還要進行員工及管理階層的再訓練。

在 1970 年代，許多商學系統都很昂貴，因此一般經理都不敢投資。近幾年來，硬體便宜許多，也促進了自動化。但是有技術的人員及合理價位的軟體仍然稀少。但有許多選擇可提供給公司有關於 GIS 的投資。有關於 GIS 的投資，在第 9 章即綜論在合理投資下，如何建立一個有效的地理資訊系統及其應所包含的因素。

### 5. GIS 未來的方向與趨勢(Future directions and trends in GIS)

過去許多 GIS 是建立應用於地方上，然而最近有強力指出使用者大多是政府部門，其正積極地建立整個國家的地理資訊系統。

地理資訊處理過程中最大問題是在測量學所產生的 sheer data volume，部分原因是地表複雜的結果，也大部分是因為地理現象很少定義的很好或很少能像汽車由電腦再複製產生。科學家們都嘗試著用不同的方法去描述相同的地方，即使受相同訓練的人也都會因不同的方法來繪製地理物件的邊界，然而其邊界常是旋繞的且非常不規則，而且地理現象也無法經濟有效地加以描述。更多的問題出自於

數值化的地理資料，不夠精密、高度的變異的或不確定的品質，如第 6 章所提的誤差。最後地理物件間的空間關係通常要求的是高度應用——特殊化(application-specific)的描述。

本書並無法解答所有問題，關於資料量(data volumes)的問題即很少觸及，但是通常誤差(error)及資料量的問題皆是由於這個世界不精確的結果。伴隨著當前概念模式之需，能夠描述自然空間變異即區格模式(compartmentalized models)及十全十美的邏輯(watertight logic)。

我們應該尋找好方法來描述這個妄想的世界以及定性考量不準確性的新方法，那是人類思考過程的完整部分，這些方向，吾人可由碎形(fractals)，模糊邏輯(fuzzy logic)和人工智慧(artificial intelligence)方法所提供的新思考著手。若只是一味的去配合技術問題的話，將導致我們目前觀念上缺憾。這個新方向隱藏在本書裏，但在這新觀念融合到原則的本體以前，仍要有相當多的研究與探討。我們應嚴肅的看待這些新的想法，可提昇我們觀念去改善地理訊系統的短處。

#### 參考文獻

- Cliff, A. D. and Ord, J. K. (1981). Spatial processes: models and applications. Pion, London.
- Hopkins, L. D. (1977). Methods for generating land suitability maps: a comparative evaluation. Am. Inst. Plan. J. October, 386-400.
- Journel, A. G. and Huijbregts, Ch. J. (1978). Mining geostatistics. Academic Press, London.
- McHarg, I. L. (1969). Design with nature. Doubleday/Natural History Press, New York.
- Vike (1981)

## 第二章 主題圖的資料結構(Data Structures for Thematic Maps)

### 1. GIS 的資料結構(Data Structures for GIS)

地理資料是複雜且包括位置、可能的位相關係和物件屬性的紀錄。地理資料以一套標準的座標系統表示其在地表的相對位置，如 UTM 座標系統，所有地理研究需要有現象(phenomenological)的觀念，如城鎮、河...，以做為分析及合成複雜的知識及建立基本區塊，可以依階層分類將其細分成不同尺度的幾個單元。如國家、縣、市、區的層級或土壤分類系統的階層或植物和動物分類系統的階層。

我們要注意的是這種地理現象依其外部形態而分，如桌子、椅子，但其真正的形式、內涵還是有疑問的，或者會隨時間改變。有關資料品質及誤差，將會在第 6 章討論。

### 2. 點、線、面：

所有地理資料都可簡化成三個基本位相觀念一點、線、面，每個地理現象能用這原理以點、線、面原理再加上標示來描述說明，例如點—樹、線—林道、河流、面—土地利用型、林型。

### 3. 地圖的定義：

地圖是一組空間位置及非空間屬性的點、線、面參照共同座標系統所組成的，通常是二維。例如樣區係由樣區位置座標(x, y)與樣區名稱表示(如圖 2-1)。

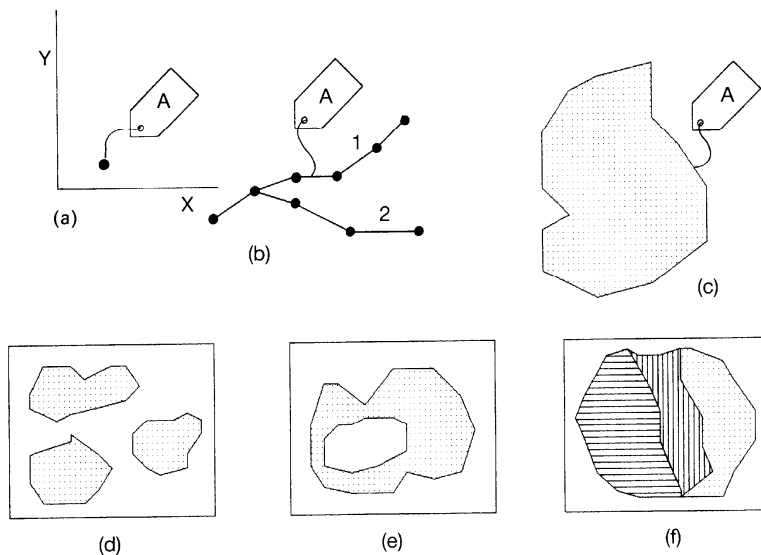


圖 2-1 地圖是一組空間位置及非空間屬性的點、線、面，參照共同座標系統所組成的(a, b, c)，地圖上某一地區屬於相同等級或有相同屬性(d, e, f)(d)為不連續區塊(e)穿孔區(f)三個不同區

圖例(Map legend)把非空間(non-spatial)屬性連接至空間實體上，非空間屬性資料由顏色、符號、陰影來表現。GIS 中非空間屬性需給予編碼及資料分析，每

一個地區(region)為多組像元(pixels)、面積和多邊形的集合，他們是由一個單獨的圖例單元(legend unit)來描述。一個地區可由許多相同的組件合成或屬不同種來描述，也可能由數值化來描述其整個屬性區塊的關係。

#### 4. 電腦內的地理資料：

電腦無法像人類一樣組織，資料必須以程式來適當描述現象、結構，資料必須儲存在磁碟內。

地理資料儲存的 4 個階段：

使用者 □ 現象... □ 資料庫... □ 硬體...

A □ B □ C □ D

A：使用者對現象構造的理解

B：現象構造以 GIS 描述

C：資料庫構造

D：硬體構造

本文僅以 A, B, C 三階段為範圍

#### 5. 資料庫結構：電腦內的資料組織(Database Structures：data organization in the computer)

在詳細討論地理資料儲存在電腦的方法之前，我們通常先考慮資料如何組織，如何儲存及輸入。雖然這對使用者並不重要，但是掌握資料的結構可幫助我們了解系統的工作方式、限制及其優點。

任何資料儲存系統必須可以擷取資料且具有快速的交叉參考能力，但是目前尚沒有一套所謂最好的儲存系統，在一般的情況下，將人力及金錢投資在有效的資料庫管理系統上，這個系統是用來控制數值資料庫輸入、輸出、儲存與擷取資料的電腦程式。

#### 6. 檔案及資料存取(Files and data access)

##### (1) 簡單表列(Simple list)：

最簡單的資料庫是把所有項目簡單地表列出來，當增加新的項目至資料庫時，是放置於檔案的末端，增加資料是很容易，但是這種擷取資料的方式很沒有效率。如一個表包括  $n$  個項目，則平均需要  $(n+1)/2$  個搜尋步驟。把資料排序或組織化是個必要的步驟，利用關鍵字查詢可以加速資料的獲取。

##### (2) 排序檔(Ordered sequential files)

數位字典中的字或電話簿的名字，係依字母次序加以建構的。增加一筆新資料時要用插入，但可以快速被找到。一般以二進位(binary)方式搜尋，即首先由中間查，再往前找  $n$  筆資料，平均需  $\log_2(n+1)$  個搜尋步驟。

##### (3) 索引檔(Indexed files)

簡單序列和排序檔(Simple sequential and ordered sequential files)需要以關鍵屬性來查詢擷取資料，但在許多應用上尤其是地理資訊個別項目(像元點、線或面)，則不僅是一個關鍵屬性，其他相關屬性的資訊都會跟過去。譬如說：有一個依土系名稱的土壤剖面序列表，然吾人希望獲得土壤深度、排水狀況、pH、



結構或沖蝕情形時，除非我們採取其他策略，否則簡單序列檔無法達成。

利用索引檔(Indexed files)去查詢原始資料檔，可以利用二種方式來增加查詢速度。(a)檔內的資料項目依照檔的主要順序，這個檔稱為直接檔(direct files)；(b)依據標題將主檔內項目的位置單獨化，成為第二個檔，稱為轉換檔(Inverted files)。就如同利用索引查詢書本內容一樣。

在直接檔內，每筆資料都包括了足夠的資訊，尋找時可跳過不需要的部份，搜尋一筆特別的資料可以簡化，並藉由建立一個簡單的目標檔，包括了主題的第一個字母及儲存位置，搜尋包括目標及資料區(data block)，平均需  $(n_1+1)/2+(n_2+1)/2$  個步驟( $n_1$ ：找 index 步數， $n_2$ ：找 data block 步數)。轉換檔索引(inverted file index)的應用，首先需要針對每一主題依序找資料，其結果與轉換檔或索引皆相似，都是獲取資料的關鍵。

索引檔可以快速進入資料庫，其缺點是：直接檔於增減記錄時，需要修正檔及索引，而轉換檔索引則將記錄放在檔的最後面，只要即時更新即可。檔案修正是很昂貴，需大量資料檔，尤其在對話環境下。另一個缺點是 indexed files 只有 key 在 index file 內，其他資訊只能用排序查詢法(sequential search method)獲取。

表 2-1 索引檔

直接檔

Index		File item
Item key	Record No.	
A	1	A <sub>1</sub>
B	n <sub>a</sub>	A <sub>2</sub>
C	n <sub>a</sub> +n <sub>b</sub> +1	:
.	.	B <sub>1</sub>
.	.	:
.	.	C <sub>1</sub>

轉換檔

Soil profile number	Attributes					
	S	pH	De	Dr	T	E
1	A	4	deep	good	sandy	—
2	B	5	shallow	good	clay	yes
3	C	6	shallow	poor	sandy	no
4	D	7	deep	good	clay	yes
5	E	4	deep	poor	clay	no
6	F	5	shallow	poor	clay	no

S=Serise, De=Depth, Dr=Drainage, T=Texture, E=Erosion.

## 索引(轉換檔)

Topic	Profiles (sequential numbers in original file)				
Deep	1			4	5
Shallow		2	3		6
Good drainage	1	2		4	
Poor drainage			3		5
Sandy	1		3		6
Clay		2		4	5
Eroded		2		4	

## 7. 資料庫結構(Data base structure)

一個資料庫包括許多檔，為了能由一個或多個不同檔快速獲取資料，必須有某種形式的結構或組織，資料庫結構主要有三種：層級式(hierarchical)、網路式(network)及關連式(relational)。

## (1)層級式資料結構(Hierarchical data structure)：

當資料有母子或1對多的關係時，層級式方法是又快又方便，其假設是每一部份都可用一個關鍵值來描述資料結構，優點是容易了解，也很容易即時反應及擴大，但很難與屬性關連。如果事前已知所有問題的結構，則層級系統方式對於資料的獲得很方便。其缺點為包含很大的索引檔(index file)，某些屬性要重覆放置，造成資料冗長，增加儲存及獲取資料的成本(如圖 2-2)。

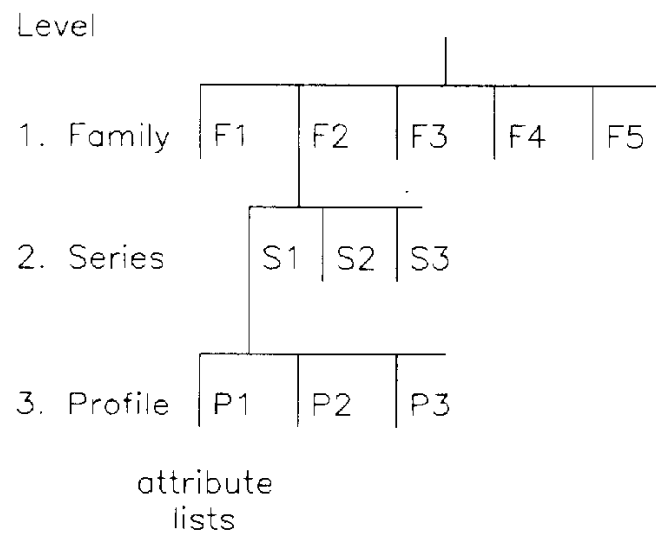


圖 2-2 一般使用於土壤科學的層級資料結構

## (2)網路系統(Network systems)：

層級系統內，執行路徑被限制在資料庫分類中。在許多情形下，資料需要更快的連結，尤其是繪製地圖時，透過座標結合各種資料庫，都接有項目(items)或要與其他圖形接合一起，網路系統可以滿足這個需要。

在繪圖上，常用含環指結構(ring pointer structure)的網路結構(network

structures)當關係或連接可以在事先指定時，網狀系統非常有用，他們避免資料重複冗長並妥善使用有用的資料，缺點是資料庫因具方向指標(overhead of the pointers)而變大，在複雜的系統中成為一個重大的部分，每次改變這些方向指標(pointers)時，則必須更新／維護，而因此方向指標結構的建立及維護成為資料庫系統的經常開支。

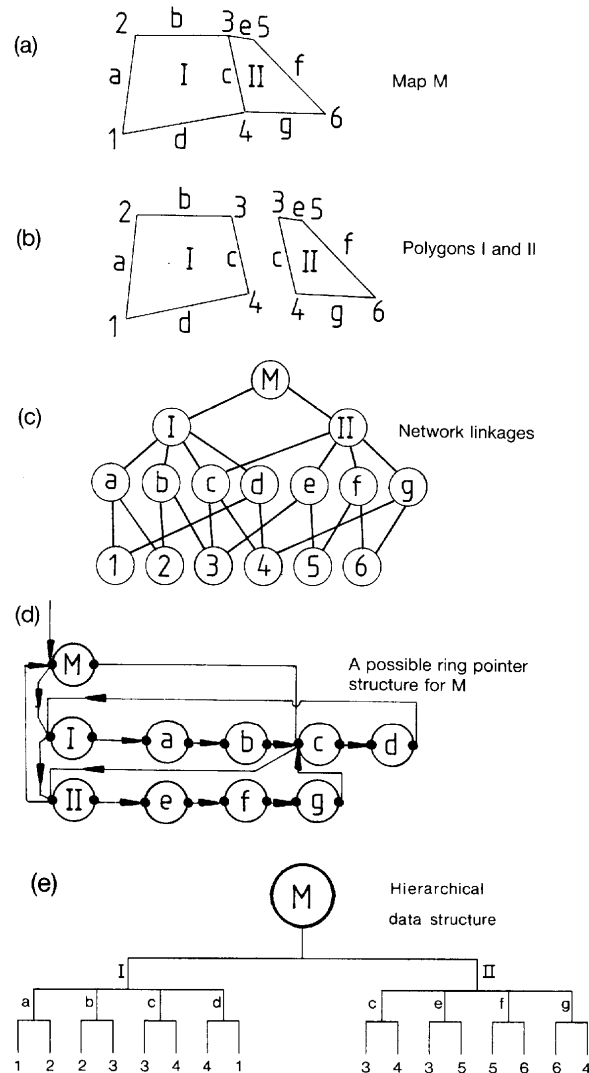


圖 2-3 簡單多邊形的網路結構(a)地圖 M，(b)兩個組成多邊形 I, II，(c)網路結構連結全部的多邊形、線、點，(d)地圖 M 可能的環指結構，(e)地圖 M 的層級資料結構

### (3) 關連式資料庫結構(Relational database structures)：

關連式資料庫結構(The relational database structures)是最簡單的儲存形式，既沒有方向指標，也沒有階層，其資料以簡單的記錄儲存稱為 tuples，包括了一組有序號的屬性值，以 2 維表格放在一起稱關連(relations)，每個表或關連是分開的檔。網路結構中的方向指標(pointers)、層級結構中的關鍵(keys)是由辨別碼(identification codes)代替，以為辨認每個檔內記錄的唯一關鍵。

使用者依據問題來定義關連，再由關連資料庫(relational database)來獲得所

需資料，其關連是不需要呈現在目前檔案中，而可以以代數關係控制程式去建立新表。

優點：構造具彈性、可解決和交集、聯集等布林邏輯(Boolean logic)及數學問題的需求；允許不同型的資料搜尋、組合及比較；增加或移除資料也很容易。缺點：有順序地尋找檔內的資料以滿足特定關係，需要很多操作，即使在快速的電腦處理，大量的資料庫也需要相當可觀的時間，同時商業的關連資料庫系統設計也非常需要技術以達合理速度，所以很貴，目前他們才剛開始應用到 GIS。

Map	<table><tr><td>M</td><td>I</td><td>II</td></tr></table>			M	I	II
M	I	II				

Polygon	I	a	b	c	d
	II	c	e	f	g

Lines	I	a	1	2
	I	b	2	3
	I	c	3	4
	I	d	4	1
	II	e	3	5
	II	f	5	6
	II	g	6	4
	II	c	4	3

圖 2-4 地圖 M 的關連式資料結構

#### (4)記錄結構(Record Structures)：

在所有資料庫結構中，資料以記錄(records)形式書寫。最簡單的記錄種類是固定長度的一維陣列，且分成許多相等的部分：固定長度記錄(Fixed length records)此種方法不方便，因各組資料數可能不同。變動長度記錄(Variable length records)每一個記錄包含一個表頭(header)為特別的屬性包含了資訊型態的訊息與所佔的空間位置(如圖 2-5)。

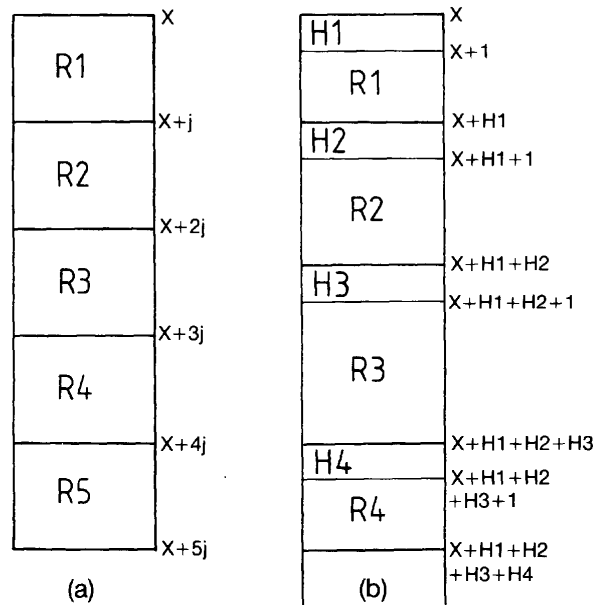


圖 2-5 二種序列表(a)固定長度的記錄(b)變動長度的記錄方式含表頭  $H_i$

#### 8. 觀念結構與地理資料的電腦表示(Perceived structure and computer representations of geographical data)

人類眼睛可以有效地辨別形狀(shapes)與形式(forms)，但是電腦則需要很正確的指引空間形態須如何操作與展示，其中有兩種相映且互補的方法來表示空間的資料屬性，一為顯性的網格式資料與另一為隱性的向量式資料，來描述空間實體。

如圖 2-6 在電腦中以顯性或隱性兩種不同方式表示椅子。

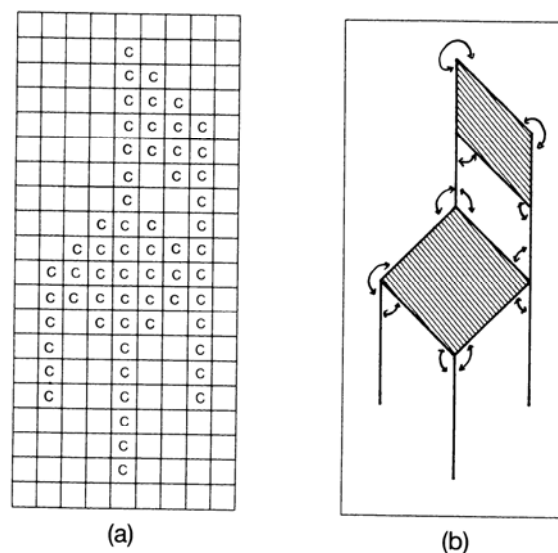


圖 2-6 椅子的影像(A)網格(raster or grid cell)(B)向量式(vector format)

網格式(raster)：顯性表示法，意即椅子是由一組在網格(grid or raster)上的點組成。所以電腦知道這是椅子不是桌子，每一個單元皆以相

同的碼，C 值表之。實際上 C 值表示一組數字或顏色或灰度...，所以椅子的資料結構可以下式表示之：椅子屬性以符號/顏色→表示各單元(cell)X，表示之。

向量式(vector)：隱性表示法亦即物體用一組由起點與終點，與某些形式的連接的線表示，這種線的起點、終點定義為向量(vector)來表示椅子的形態，線間的指標，提示電腦，線如何連結一起形成椅子。

這個資料結構為：物體屬性→一組向量(set of vectors)→連結(connectivity)

二者之差別：

- 1.儲存空間：隱性表示法(implicit representation)需要較少的數目，亦即較少的儲存空間。(如圖 2-6，向量式用了 11 組 xy 和 14 個連結指標，網格式則用去了 60 個單元。
- 2.美學上：向量表現比網格式表現來得漂亮。
- 3.資料更新上：當連接資訊時，網格式允許直接做空間探討。當物體形狀及大小改變時，網格式表示法比向量式表示法來得快速容易，網格式(raster)只要馬上消掉舊數值換上新的就可以了；向量式(vector)則不只要座標還要重建其連結性。

吾人知道至少有兩種基本方式可以表示位相資料(topological data)，其綜論如下：

網格式表示法(raster representation)：由座標來定位的一組單元(cell)，每個單元(cell)與其屬性的值是互相獨立。

向量表示法(vector representation)：三個主要的地理要素為點、線、面，除了他們未覆蓋面積外點是與單元(cell)相似，線和面是一組互相連接的座標組，可以與其屬性連結。

要在意的是在地理資料庫的網格式與向量結構之間連結及使用網格式或向量結構做資料展示，並不是唯一或必須的方式，雖然常有這種情形。譬如說：現代對談式電腦輔助設計和製圖系統雖以向量結構資料庫表示，但都是使用顏色網格式展示和向量繪圖機(詳第 4 章)。

## 9.網格式資料結構(Raster data structure)：

最簡單的網格式資料結構(raster data structure)包括了網格式單元(grid cells)有時稱作像元(pixels)或圖素(picture elements)。每個網格式單元(grid cell)由一個列(row)、行(column)決定，以描述繪圖屬性的形態或數值。在網格式結構裏，一個點(point)是由單一網格式單元表示之；一條線是由許多臨近的單元串連配合一個方向而成；一個面(area)則由許多臨近單元組合而成。此種資料結構，因為用一序列的行、列貯存操作與展示，易於由電腦作業，尤其以 FORTRAN 寫的程式。這個表示地理資料結構的二維表格所描述的為不連續的，但是量化的，這對估計

長度和面積有重要影響。譬如，圖 2-7(a)表示 a 與 c 的距離為 5 單元，2-7(b)a 與 c 的距離，依計算單元邊緣，還是計算整合單元，是可為 7 或 4 單元。

1.a~c 的距離計算

圖(a)為 5 個單位長

圖(b)為 4 或 7 個單位長

2.面積計算

圖(a)為  $3 \times 4 / 2 = 6$

圖(b)為 7 個平方單位

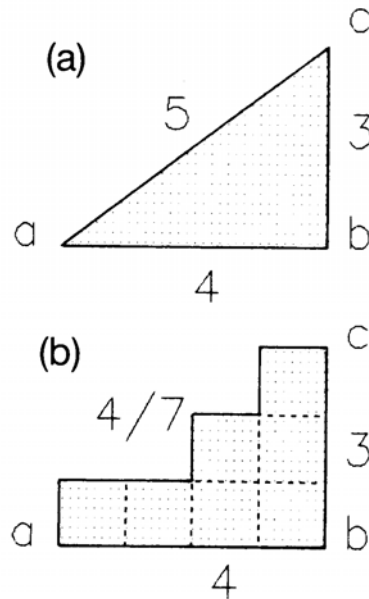


圖 2-7 因為 cell 量化的效應，網格編碼會影響距離和面積推估，

網格表示法為假設地理空間是平坦的笛卡兒(Cartesian)表面，解析度或網格資料的尺度是資料庫的網格(cell)大小和地面的網格大小間的關係。

10.地圖疊合(Map overlays)：

因為在二維空間的陣列，每個網格只能擁有一個碼，不同的地理屬性則由不同的笛卡爾陣列(Cartesian array)來描述，稱之為疊合(overlays)，如圖 2-8。疊合觀念由推估二維度序列，若三維度者則由圖 2-9 表示疊合網格資料結構。

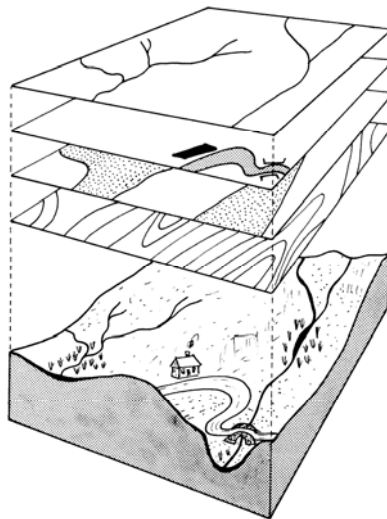


圖 2-8 疊合的觀念：真實的世界是由一序列的主題圖圖層分別記錄各主題資料 (如地形、土壤型、道路、河流等...)

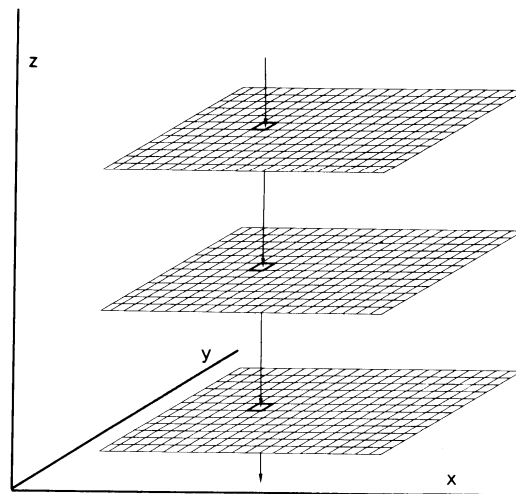


圖 2-9 三維陣列使用於網格資料庫結構套疊的編碼製圖

### 11. 套疊結構內的參考像元(Referencing pixel in raster overlay structures)：

資料如何在電腦中組織起來，且適當地獲得資料，並減少儲存及需求的程序。假設在資料庫中(資料值間一對一的關係、位置、像元)每個網格在每次套疊中是一個獨立單元，則可能組成三種基本且相當的方法(如圖 2-10)。

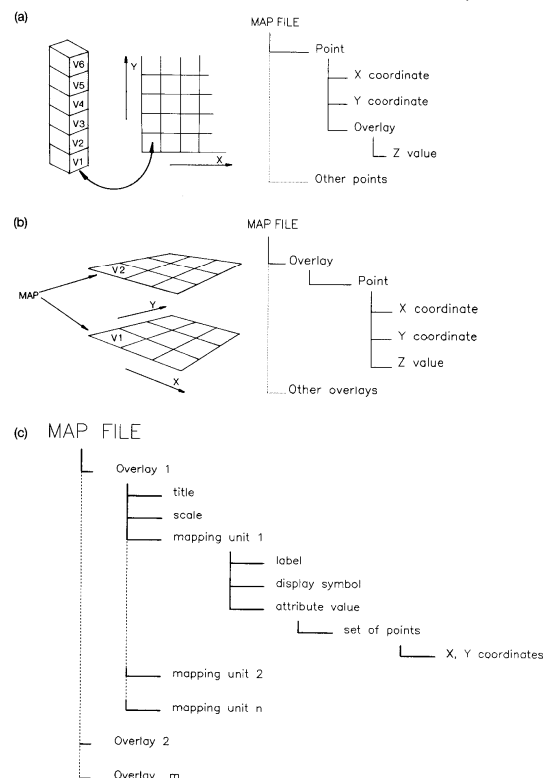


圖 2-10 三種網格資料庫結構

a：每個網格被直接地參考

b：每個套疊被直接地參考

c：每個製圖單元或地區被直接地參考



## 12. 儲存網格資料的壓縮方法(Compact methods for storing raster data)：

當每一個網格有單一的值時，每個套疊編碼共需  $n$  行  $\times m$  列  $\times 3$  ( $x, y$  座標及屬性值) 個值，有四個主要方法可壓縮存檔：

(1) 連結編碼(Chain codes)：(如圖 2-11)區域的邊界用單位向量表示方向(東=0、北=1、西=2、南=3)，為  $0, 1, 0^2, 3, 0^2, 1, 0, 3, 0, 1, 0^3, 3^2, 2, 3^3, 0^2, 1, 0^5, 3^2, 2^2, 3, 2^3, 3, 2^3, 1, 2^2, 1, 2^2, 1, 2^2, 1, 2^2, 1^3$  (如  $0^3$ =向東三格； $2^3$ =向西三格)。

優點：可提供描述區域的壓縮貯存方式並做其他運算，如估計面積及周長或很容易偵測出是轉彎或凹陷的形狀。

缺點：1. 沒有全格的描述，於執行套疊一個單位或一段是很困難的。

2. 二個區域邊界會被儲存 2 次，資料較冗長。

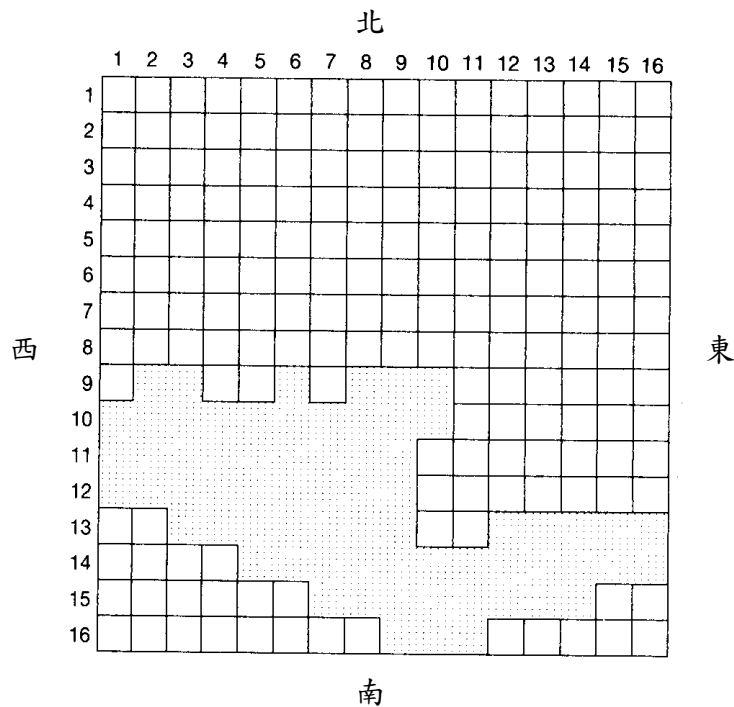


圖 2-11 以網格式地圖表示之簡單區域

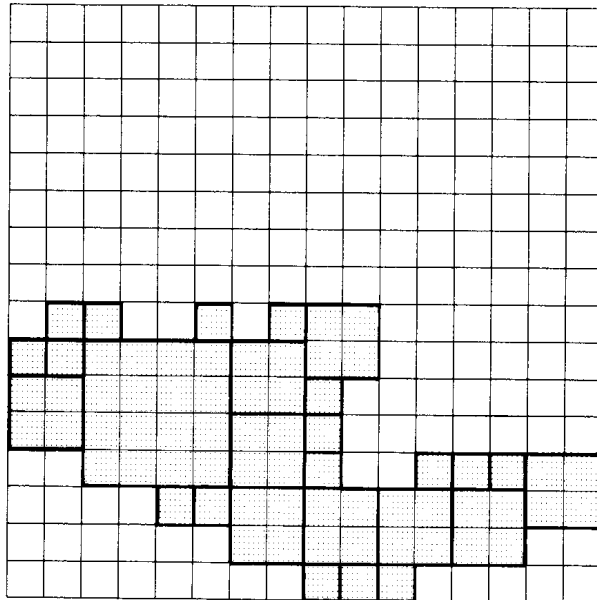
(2) 行長編碼(Run-length codes)：以行(row)為單元，由左至右記錄儲存於圖上的起始網格單元。

如：Row 9    2, 3    6, 6    8, 10  
 Row 10    1, 10  
 Row 11    1, 9  
 Row 12    1, 9  
 Row 13    3, 9    12, 16  
 Row 14    5, 16  
 Row 15    7, 14  
 Row 16    9, 11

優點：儲存上較改進，也可減少資料儲成簡單網格資料庫的體積。

缺點：資料適度壓縮，會增加繪圖時轉換的過程。

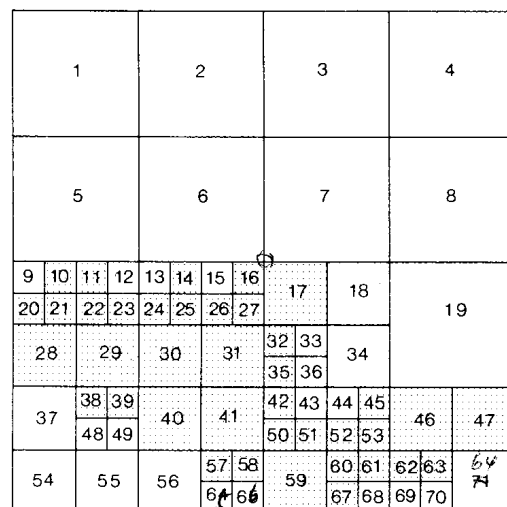
- (3)區塊編碼(Block codes)：用正方形來畫區域，每筆資料包含三數；原點(中間或左下方)、每個方格的半徑，這就稱為中央軸轉換格式，(Rosenfeld 1980)在區域中可以畫的正方形越大邊界就越簡單，就越有效。



$$\begin{array}{rcl}
 1 \text{ 格} \times 17 & = & 17 \\
 4 \text{ 格} \times 9 & = & 36 \\
 16 \text{ 格} \times 1 & = & 16 \\
 \hline
 & & 69
 \end{array}$$

圖 2-12 由中間軸轉換區碼(medial axis transformation block coding)描述的簡單區域

- (4)四分樹(Quadtrees)：把四方形連續四等份細分，再看那一象限全部包括在區域內  $2n \times 2n$  點的樹狀圖，根節點(root node)最多有  $n$  層高度，每個節點有 4 個分枝代表 NW, NE, SW, SE, 4 個象限，每個 node 由 2 bites 來表現，如圖 2-13 所示，此區域結構可由四分樹表示謂之四分樹(Quadtree)。於圖 2-11 內表現圖 2-14 的結構， $\uparrow\uparrow$ -terminator in,  $\downarrow\downarrow$ -terminator out,  $\uparrow\downarrow$ -node in,  $\downarrow\uparrow$ -node out。四分樹是變數解析度(variable resolution)序列，在不須額外儲存缺乏之詳細部分時，可表現詳細資料，如圖 2-15(a-d)與圖 2-16。



$$2^4 \times 2^4 = 256$$

圖 2-13 由四分樹表示之簡單地區

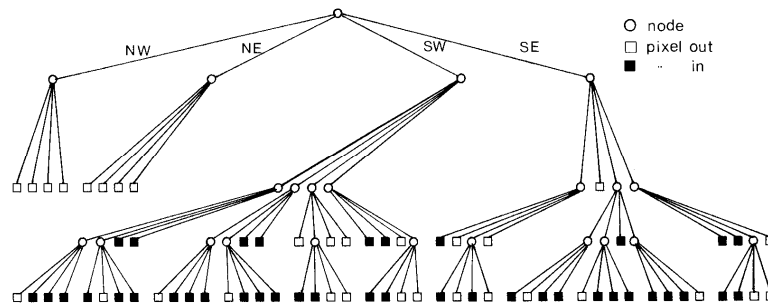


圖 2-14 圖 2-11 的簡單區域之四分樹結構

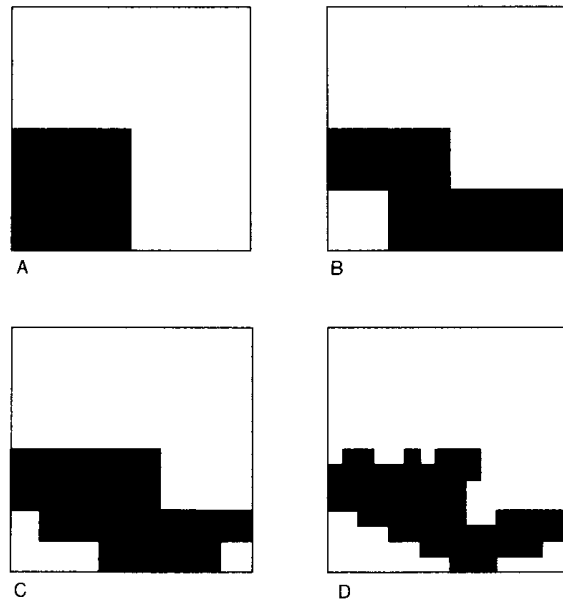


圖 2-15 在四分樹層級式的四個附層內視覺的表現

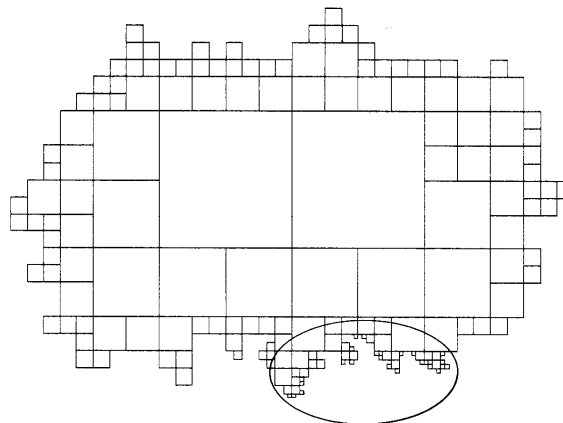


圖 2-16 由四分樹結構分割區域在橢圓區內詳細表示以四為基礎之分枝的 6 個層級

優點：標準的區域性質可容易及有效的計算，解析也可變化。

缺點：樹狀圖不是不會改變的，相同形狀及大小的區域，可能有不同的四分樹。

總結：網格資料結構：

行長編碼：當像元很大時最有效。

區塊編碼：當解析改善，像元數目增加時有效。

四分樹：當解析改善，像元數目增加時有效，多種解析度時更有效。

### 13.地理實體的向量資料結構(Vector data structures for geographical entities)：

本節敘述一個範圍的向量結構用在 GIS 來儲存點、線、面資料。

(1)點：為單獨的 x, y 座標對，及其他資訊，如顯示的大小及表示的方向。

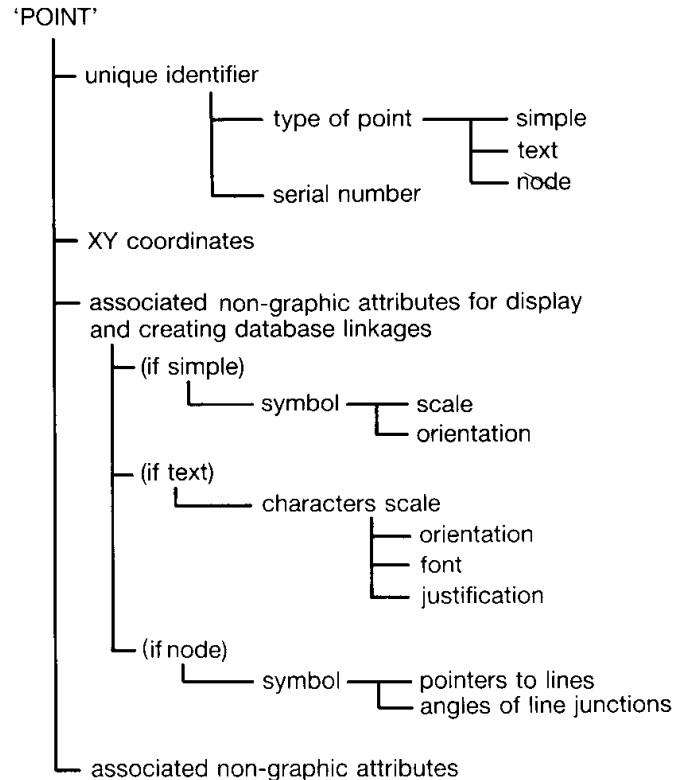


圖 2-17 簡單點實體的向量資料結構

(2)線：最簡單的線包括 2 對 x, y 座標（起點及終點）加上顯示符號的記錄。弧、鏈線是一組 n 對 x, y 座標構成的連續複雜線，每截線段越短，x, y 座標對的數目越大，鏈越接近於複雜的曲線。

(3)網路：線及鏈(chains)未具空間資訊關連性，利用節點(nodes)接上，形成網格節點，還有角度可調整，這種簡單的連接，因每個座標記錄了 (n×chains+1) 次會造成資料冗長。

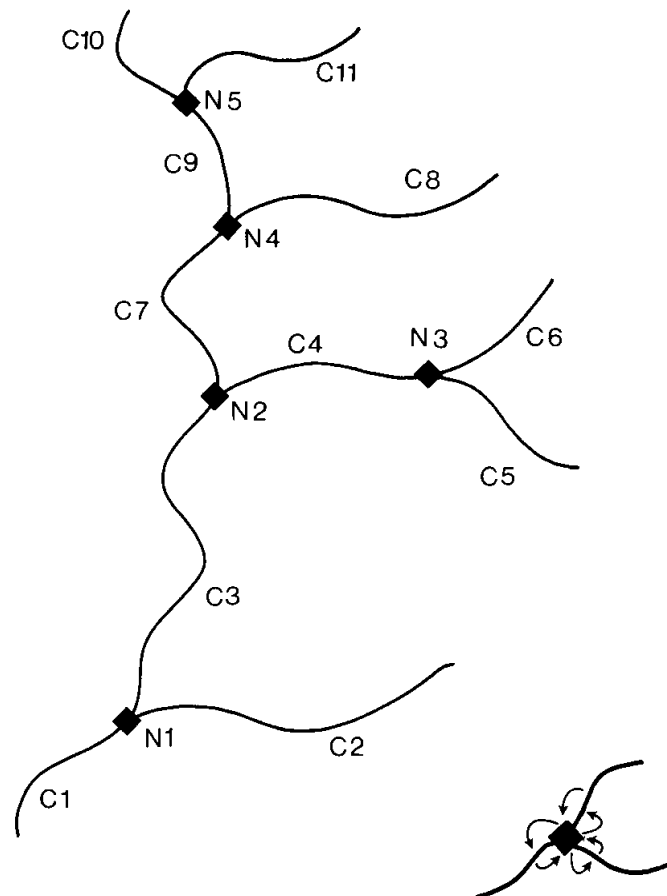


圖 2-18 使用節點以攜帶連結資訊的線網路的向量資料結構

(4)面：在向量資料庫中，多邊形的面積（稱為區域）有時可以用不同方式表示多邊形資料，而資料的建立目的在描述面積的位相性質（如：形狀、鄰近區域、位階），如同主題圖資料般的顯示這些基本空間區塊的相關性質。

地理資料產生之多邊形網路(polygon networks)的需求：1.圖上的每個地區(region)有其唯一的形狀、周長及面積 2.地理資料分析時，鄰近多邊形(polygon)資料需以相同方式記錄 3.主題圖的多邊形(polygon)並不是都在相同的土地上。

(5)簡單多邊形(Simple polygon)：描述多邊形最簡單的方式就是簡單鏈(simple chain)的擴大。

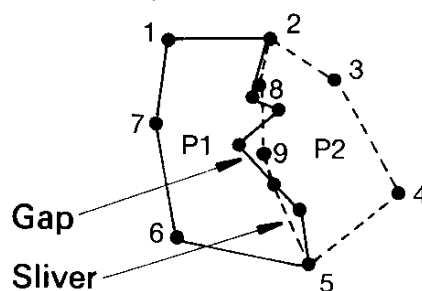


圖 2-19 簡單多邊形結構的缺點在於兩個多邊形的邊界線需數化與儲存兩次，並形成"裂片(slivers)"和"孔隙(gaps)"的位相誤差

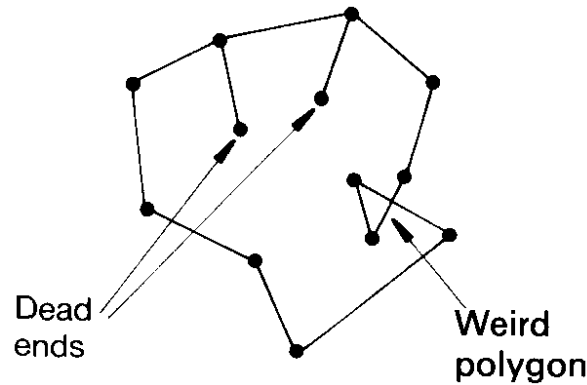


圖 2-20 簡單多邊形邊緣內的位相誤差，其不完全的連結(死巷, dead ends)和不容許的環形位相(曲折的多邊形, weird polygon)必須從資料上移除

#### 向量資料結構

優點：簡單

- 缺點：1.相鄰多邊形的線要數位化及儲存 2 次，會造成誤差(如圖 2-19 和圖 2-20)；  
 2.沒有鄰近區域的資訊(如位相關係)；  
 3.除非純粹地理構造，不可能有孤島的存在；  
 4.如果地形邊界是正確的或它是不完全的死巷(dead-end)或曲折的多邊形(weird polygons)，至目前並沒有簡單的方法可以檢查。

簡單的多邊形結構可以擴大，但是並不能避免基本問題。

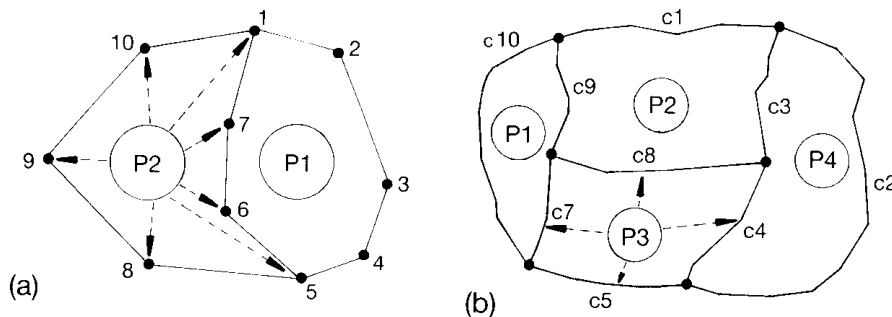


圖 2-21 (a)多邊形資料結構內所有的多邊形座標直接從多邊形紀錄上參考。  
 (b)與(a)地圖相同，但所有鏈直接從多邊形紀錄上參考

#### (6)點目錄的多邊形(Polygons with point dictionaries)：

每個多邊形相連結之所大點的座標依序字典排編號列出參考，其優點是相鄰多邊形的邊界是唯一的，但鄰近區域函數的問題依然存在，且結構也不容易允許鄰近多邊形的邊界被刪除或溶掉，如果重新編號或分類以改成為相同等級，則島狀多邊形的問題仍然存在。鏈目錄(chain dictionary)之優點：定鏈的界限可由其運算清除減化其大小。

#### (7)外顯多邊形的位相結構(Polygon system with explicit topological structures)：

孤島及鄰近者只能在複合的外顯位相資料結構中才能正確掌握，位相結構能以二種方式來建立，其一，在資料輸入時建立位相連結，或是使用軟體從一組連

環、建立位相。前者負荷在操作時推入建立位相，後者則依賴計算能力。二者都會增加資料的儲存量。

最早建立於地理資料結構之外顯的位相關係的是美國普查局的 DIME(Dual Independent Map Encoding)系統，其資料檔基本組成是由簡單線段之兩端定義，複雜的線由一連串的線段呈現。

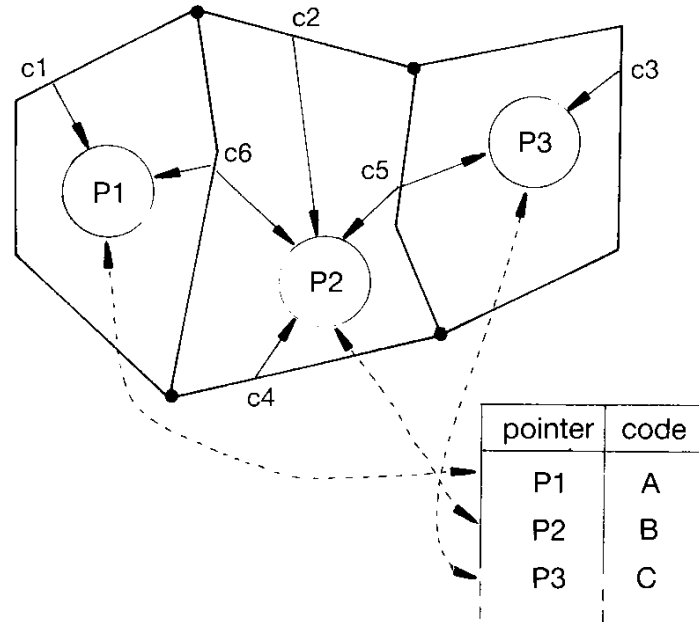


圖 2-22 多邊形資料結構內所有多邊形是從鏈(弧或邊界線)參考

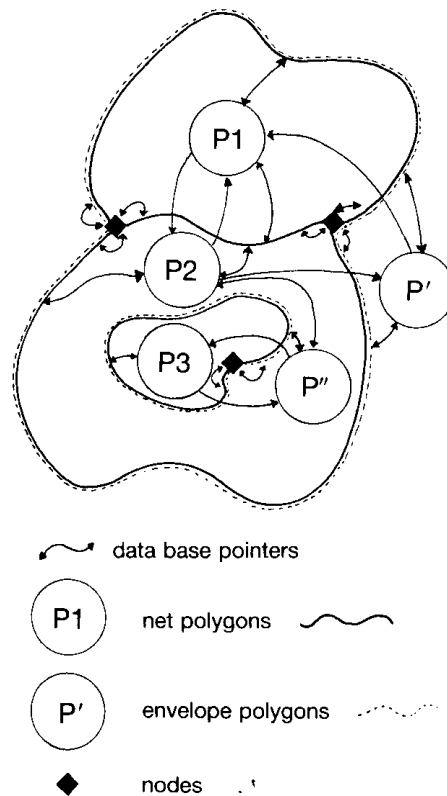


圖 2-23 多邊形的豐富位相網路圖

荷蘭土壤調查研究中心發展出一個簡單、有效的方法，可把多邊形資料以微電腦處理。每個鏈是以列出 x, y 座標對及相鄰地圖區域的二點存入。多邊形的名字被分別儲存在包含相同指標的表格中。包含島多邊形的面積由一個點內多邊形(point-in-polygons)尋找算術而計算，最好的方法是由完全的位相結構資料庫來建立。

(8)多邊形網路結構完全地位相關係(A fully topological polygon network structure)

本節敘述如何由一組邊界 chains 或 strings 來建立圖 2-23。

建立資料結構的五個步驟：

步驟 1.連結鏈至網路邊界：

鏈首先依 x, y 座標依最小最大法分類，在尋找鄰接的鏈時可節省時間，然後檢查那些鏈他們相交連接點建立起來在 chains 的來端 chain data record 擴大包括 chain pointer 及角度。

步驟 2.檢查多邊形的封閉性(checking polygon for closure)：

網路可容易的檢查其封閉性藉由掃描修節鏈記錄(modified chain records)觀察是否有環指(pointers)去接另一個鏈(chains)。

步驟 3.把線連成多邊形：

第一步是由地圖外邊界產生一個新的封套的多邊形，這個封套實體(envelope entity)包含的記錄包括了(a)單一的確認者(b)確認外框多邊形的編碼(c)一個環形指標(d)一張指標對邊緣鏈的清單表(e)他的面積(f)它的範圍。

封套多邊形(envelope polygon)使用者看不到其唯一目的，在建立網路之位向構造，envelope polygon 之產生是 chains 沿著外圍，以順時針方向並選擇每個 junction 的最左邊鏈(chain)，每個鏈的唯一 identifier 與其他資料一起被記錄並儲存，並設定一個旗幟表示每個鏈被橫過一次 outer—同樣的地方開始，順時針找在每個 junction 選最右邊的鏈，記錄鏈被橫過的次數，橫過 2 次就不找了，回到起起點。

如同 envelope polygon，每個多邊形實體可以取之幾組資訊。

(a)唯一的確認者(b)普通的多邊形編碼(c)來自外邊框多邊形的環形指標(d)所有邊界鏈的列表(e)網路內鄰近多邊形的環形指標(f)長方形邊最小和最大的 x, y 座標(範圍)。

搜尋進行至同一階層相同網路內的下一個多邊形，直到每個 polygon 都建立，當最後一個多邊形找完，環形指標回到封套多邊形可確定每個邊緣皆與 2 個多邊形連接。

步驟 4.計算多邊形面積：

用梯形定律計算個別多邊形的面積，先計算全部再扣除表面的孤島，然後當作屬性存起來。

步驟 5.關連多邊形的非地理屬性：

首先是把每個 polygon area 內唯一的本文數化，其次是在每個 polygon 中間



用電腦寫下多邊形的單一確認者，同時印出每個確認者。

由以上 5 點，資料結構有以下優點：

- (a)多邊形網路完成整合而沒有孔隙、裂片及其他額外的座標。
- (b)全部多邊形、鏈及關連屬性都是互相連結單位的一部分，因此各種鄰近區域分析是可能的。
- (c)大陸—島嶼的數目未受限制。
- (d)資料庫的區域正確性只受到數位及電腦字數化和長度的正確性影響。
- (e)資料結構對於資料收集和輸入有很少的含意。

(9)點內多邊形的搜尋(Point-polygon search)：

圖 2-24 比較點與 polygon 範圍座標可知 點在圖內或外

圖 2-25 水平直線穿越測試點，如果與 polygon 截了奇數段所以 點在多邊形內檢查島 P'在 p 內部，首先把 P'分成一些水平 bands，bands 的起終點如上測試如果是奇數所 P'完全被包括。

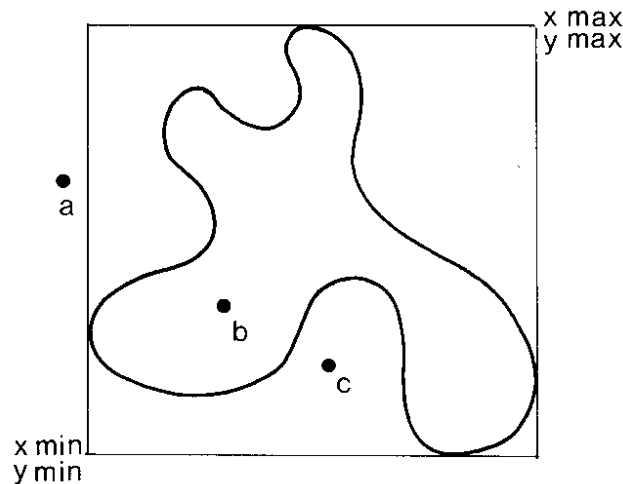


圖 2-24 解決點內多邊形問題，點 a 因為位於範圍外是容易除去，點 b 和點 c 須再處理

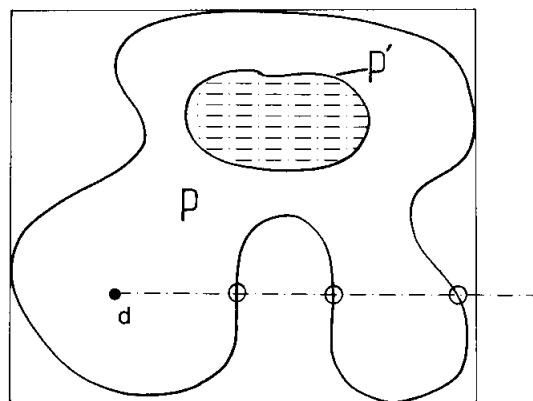


圖 2-25 假設從點 d 上畫條水平線在多邊形邊緣上如為奇數交叉點，則點在範圍內。

## (10)其他向量資料庫結構的圖徵(Other features of vector database structures):

圖層(Layers):基本上圖層的數目是無限的,只受到儲存空間的限制,不像網格系統在電腦輔助設計(CAD)向量系統中,overlay/layer 用來分割空間實體的等級,其主要目的是畫圖及展示 list and sequential senches:在每個情況下,基本的實體點、線、多邊形都被一個指標(pointer)或標籤(label)唯一的辨認,由這些標籤來討論位相關係,這些標籤有些稱為集合索引指標(master index pointers),這一系列的表(list)是進入其他資料庫的關鍵,這個表有 2 個主要問題:1.它很少是完全連續的,在編輯時,會產生孔隙或增加長度 2.搜尋的時間隨表格長度遽增。

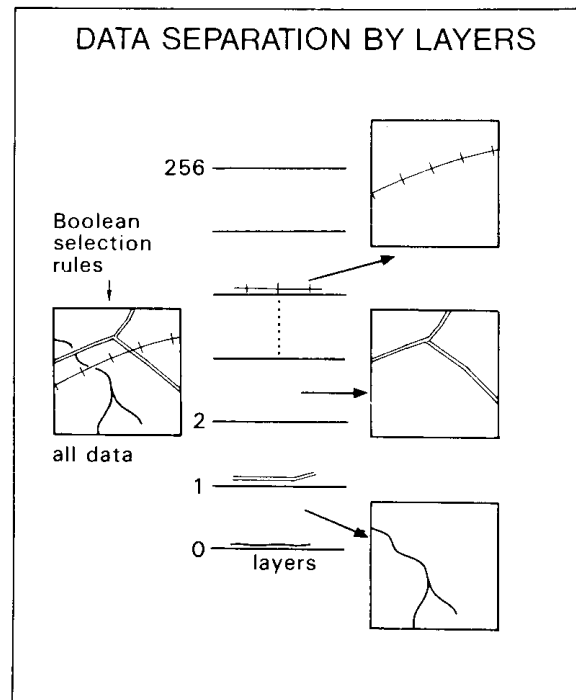


圖 2-26 視覺化"圖層"結構的圖解,使用於許多向量製圖系統,以分離成不同主題圖,每個圖層再使用於分離成一個主題圖,不同地圖可由單一視覺化圖層建構或予以組合,數字為各圖層的相對位置。

有二個基本上不同的方法去處理資料庫太大的問題。一個是強力計算快速掃描指標陣列(pointer arrays),或是把 master index array 濃縮成小的鄰接的區域。

## (11)表和序列查詢(lists and sequential searches)

到目前為止,所有向量結構所描述的只有關於點、鏈和多邊形的關係所構成主題圖,大家知曉,每件基本實體一點、線、多邊形皆是由唯一指標或標籤來辨認。位相結構只可以由這些標籤來轉換。這些標籤有時稱做主要索引指標(master index pointer),通常在一序列表用一關鍵項去獲得其他資料庫如圖 2-27(a),其他有關主要索引指標結構的方法不僅依實體型態,而且也依空間位置決定如圖 2-27(b)。

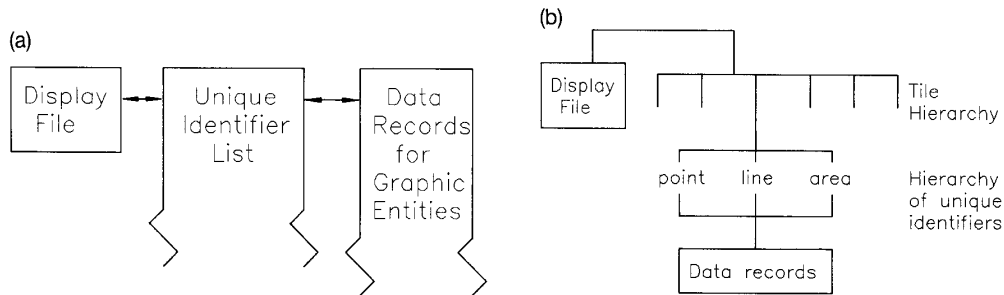


圖 2-27 (a)使用於向量圖形之簡單序 (b)為較佳的設計，係將資料庫分類成磚瓦(分別的區域)和地理實體

#### 14.主題圖資料結構—網格與向量的選擇(Data structure for thematic map—the choice between raster and vector)

在以前網格及向量二者是不相容的，因為 raster 需要大量的電腦記憶體儲存及由向量結構(vector structure)獲得的空間解析層處理影像，過去使用者被迫選擇使用網格式是允許簡單空間分析但造成不美的地圖。向量式則提供了資料庫可管理的大小及圖形優美，但對空間分析非常困難。

近幾年來，才明瞭什麼是重要的觀念問題，就是技術問題。電腦由卡片打孔進步至彩色螢幕。畫圖的品質並非唯一的技術限制。早期向量過程技術進步只是因為向量是表現畫圖最受歡迎的。在某些情況下，網格較有效率，如計算周長、面積、平均、總和…等，在視窗(Windows)環境下：拖曳及截取時，網格比向量容易，連接網路只有向量才可行，最近向量操作配合快速運算數字發展，能解決一些空間問題。

利用壓縮網格資料結構，可減少貯存的體積，反之自覺網路結構會增加資料庫大小。

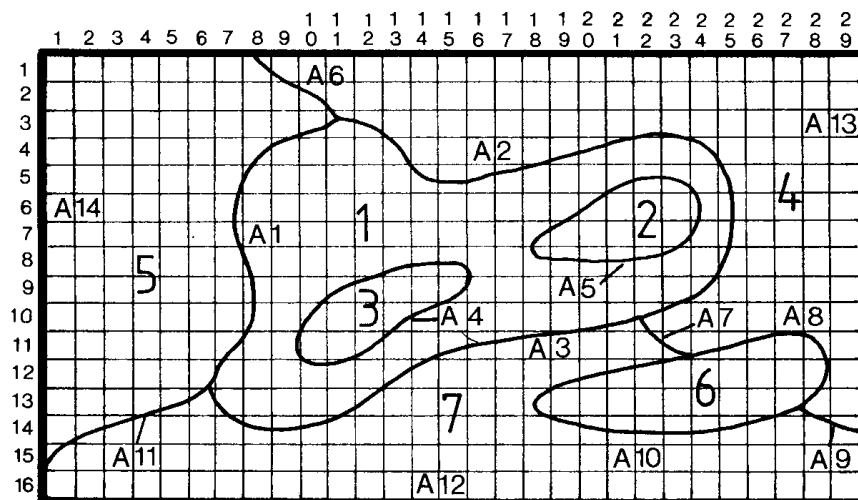


圖 2-28 一個簡單地圖可用向量式或網格式描繪

多邊形  $P_1$  是多邊形網路的一部分，稱為"土壤圖"。 $P_1$  包含二個多邊形  $P_2$ 、 $P_3$ ，且由  $P_4$ 、 $P_5$ 、 $P_7$  多邊形所圍繞，有  $A+n$  個屬性與  $P_1$  所描述空間的性質。

現在網格或向量的問題已經不存在，因為二者皆為有效的方法，且二者之結構是可互相轉換的。vector 轉換成 raster 很簡單，也可自動完成 raster→vector 複雜過程，惟要減少座標數以 weeding 的過程。

最好空間資料能以 raster 及 vector 二種形式表現，尤其是線或邊界需以網路連接或以特別型態畫出，且二者之間的空間需以某種記號或顏色填滿。

表 2-2 向量式與網格式的比較

<u>Vector</u>	<u>Raster</u>
優點： <ul style="list-style-type: none"> <li>• 描述現象佳</li> <li>• 壓縮資料結構</li> <li>• 網路連結可完全描述位相關係</li> <li>• 繪圖正確</li> <li>• 可截取、更新、圖、屬性</li> </ul>	優點： <ul style="list-style-type: none"> <li>• 資料結構簡單</li> <li>• 疊整合及圖形結合容易</li> <li>• 易於不同的空間分析</li> <li>• 模擬容易</li> <li>• 技術便宜</li> </ul>
缺點： <ul style="list-style-type: none"> <li>• 資料複雜</li> <li>• vector polygon maps 結合或與網格 overlay 困難</li> <li>• 因為每個單位有不同的形式，所以模擬困難</li> <li>• 展示繪圖、昂貴，尤其在高品質色彩和陰影時</li> <li>• 技術昂貴，尤其是複雜的軟體和硬體</li> <li>• 多邊形內做和過濾是空間分析不可能</li> </ul>	缺點： <ul style="list-style-type: none"> <li>• 繪圖資料體積大</li> <li>• 如加大 cell 減少資料體積會造成圖形失真，而</li> <li>• 粗的 raster map 比用細線畫的不漂亮</li> <li>• 網路連結建立困難</li> <li>• 投影轉換耗時</li> </ul>

在向量標誌法，多邊形的邊界項目類型是由相同方法設置。項目資料關係包含邊界的屬性，即建立弧形邊界清單和其內部洞穴邊界的邊界項目清單。弧形的資料結構包括多邊形順著邊界的左、右方向屬的資料結構和他們的座標清單。

為了描繪多邊形內  $P_i$  的網格標誌法，他只須改變從邊界到像元的空間描繪的輸入記錄。

多邊形  $P_i$  能完全由網格式或向量式描繪。在最高層，多邊形被一組相關"項目名稱"、"項目類型和項目資料"列表所描述，然而名稱和類型是簡單的記錄，"項目資料"是一組包含"屬性、空間關連範圍和空間描繪"的關係。

我們可以預期的是地理資訊系統使向量式和網格式資料結構相輔相成，其重要性是提昇對高解析度、壓縮資料結構的需求，並提昇資料分析的操縱能力。

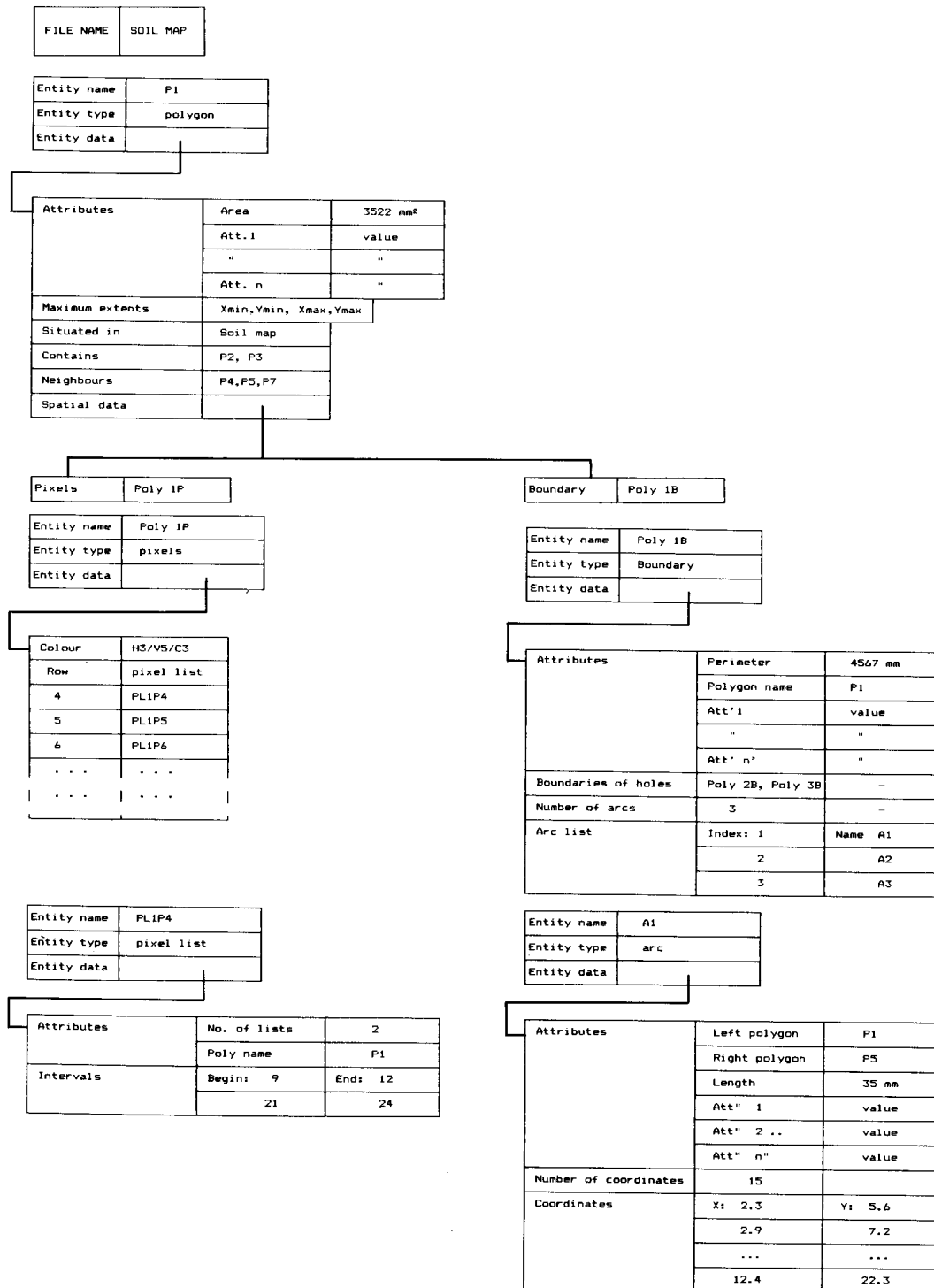


圖 2-29 結合向量與網格表示之可能關連資料庫結構

圖 2-29 表示多邊形 P1 可以完全由網格或向量表示之，在最高層級多邊形係由簡單關係組來描述，其內包括實體名字、實體的型態和實體的資料——資料包括屬性、空間關係、內容和空間描述)。

### 第三章 數位高層模型(Digital Elevation Models)

地誌圖(choropleth map model)不能解決連續性變化表面的狀況，需用等值線。等值線適合展示連續性表面改變，但不適合數值分析或模型化。空間連續變化的數值描述要用數值高程模型(digital elevation model, DEM)、數位地型模型(digital terrain model, DTM)。

#### 一、DEM 的需求方向：

- 1.在國家資料庫中儲存數值化地圖的高層資料
  - 2.在道路設計、土木工程、軍事工程課題、填挖方(Cut-and-fill)問題
  - 3.三度空間展示土地形態
  - 4.視域(軍事、景觀規劃)分析
  - 5.計畫道路、水壩位置
  - 6.統計分析及比較不同種類的地形
  - 7.為了計算坡度圖(aspect map)用來準備的陰影坡度圖(shaded relief maps)，協助地理形態研究或計算沖蝕和逕流。
  - 8.當作展示主題的背景或結合地面高低起伏的資料，以便表現主題資料——如土壤、土地利用型、植生圖
  - 9.提供資料以完成地景(landscape)的影像模擬模式(image simulation models)
  - 10.由其他連續性、變化的屬性取代高度 DEM 可表現表面上的旅行時間、花費、人口、視覺美的指標、污染程度、地下水位等等
- 本章目的：DEM 可用到的不同敘述的方法，如何獲得資料，可獲得的結果種類

#### 二、表現 DEM 的方法(Methods of representing DEMs)

表 3-1 地形表現方式

##### A.數學方式

A.1 全球—傅立葉序列  
多項式

A.2 地區—規則區塊  
不規則區塊

##### B.影像方式

##### B.1 點資料

規則—單一密度  
變動密度

不規則—三角網

邊界特徵法—頂點、谷點、途徑、轉折界

##### B.2 線資料

橫切面描述

縱切面描述

特徵線—脊線、谷線、海岸線、崖線

(1)數學區塊方法(mathematical patch methods)：依據連續三維函數的數、高度平滑化。

local method：把表面分成正方形小格或面積近似之不規則小塊，在小塊中觀測到的點來配合表面，在小塊的邊緣會用到加權的方式來確認表面契合。雖然世界線上表面的坡度不是連續。Piecewise approximation 在地形圖上並不討喜，但普遍用在複雜表面模型，近來更於用插入地下水、土壤性質或其他環境資料之變化與其他方法比較，將在 chap8 比較之。



圖 3-1 以區塊方式繪製之等高線圖可以清楚看見資料缺少的區塊，而加以增補

### 三、影像方法(Image methods)：

線模式(Line model)最普通的是等高線，常用自動掃描而得到等高線圖，數化的等高線並不特別適合計算坡度或製作影像坡度圖(shaded relief models)，所以

它們通常轉換成點式(point model)，像不連續高度矩陣(discrete altitude matrix)。Osword & Raetzsch (1984)制訂一個系統產生個別的高度矩陣，由一組多邊形，以人工或網格掃描數化等高線稱為 Graz Terrsin Model 連接上 SYS SCAN 把數化等高線擴大為高度矩陣。

The Graz Terrain Model—以幾種方式產生高度矩陣，首先，大小相近的格子蓋在等高線多邊形和橋和排水線的數位影像上，所有格子蓋在等高線上或立即接著的標示高度值，其他 cells 標示值為-1，以下列步驟標出高度值，是以線性插入狀資料庫視窗，插入是以 4 個出向搜尋線 N-S, E-W, NW-SW, NW-SE 進行，比較已知高度值的 cell 高度差，計算區或內最陡的坡度，在每個視窗中，坡度分成四級由最陡的坡度開始把未定高度的 cell 設定高度重覆這個過程，除了平坦地區以外平坦地在 DEM 中陡的部分計算完工後才個別計算。

Point model：

(a)高程矩陣(altitude matrices)：因為矩陣(matrices)可以電腦掌控，尤其是在網格 GIS 內，高度矩陣為 DEM 的最有用形式，英美可完全由粗略(course matrices)覆蓋(grid use 63.5m for U.S.A.)由 1：50000 或 1：25000 地圖空照圖越來越有用。其缺點：(a)在同一地區大量資料冗長(b)不改變格子大小，無法調整地區不同信賴複雜度(c)在某些計算上(如視覺計算)沿網格軸會有擴大情形。

指標上資料太冗長的問題，可由高度大部分解決，立體照片圖可自動掃描網格內並增加面積內複雜性的信賴度。貯存資料部分太冗長仍是問題。沿格線的平行線計算軸的位向可減少簡單的行列搜尋，其他角度仍需要三角計算正確角度和距離。

除了這些缺點外，高程矩陣(altitude matrices)是 DEM 最易獲得的形式。

不規則空間點樣品可以二種方法產生 DEM(1)把規則網格疊合在點上，再以內指導出高度矩陣，內插也可產生較細的矩陣。(2)用這些不規則化空間的點的資料做為三角化系統

(b)不規則三角網(The Triangulated Irregular Network, TIN)：係由 Peucker 和其合作的設計的系統可避免高程矩陣(altitude matrix)冗長，在計算上比只有數等高線者更有效率。TIN 是個 terrain model 用連續連接的在不規則空間節點或觀測點三角化。TIN 可以只收集額外的資訊在複雜的起伏(relief)面積中而不需收集冗長的簡單起伏面積，同時資料捕捉也可由山脊、流線及其他能被數化的重要地形特徵而達到所需的正確度。TIN 是向量結構、連接點由一個 node 至鄰接 node 由順時針方向連接而自北方開區域外部以 dummy node 表示。



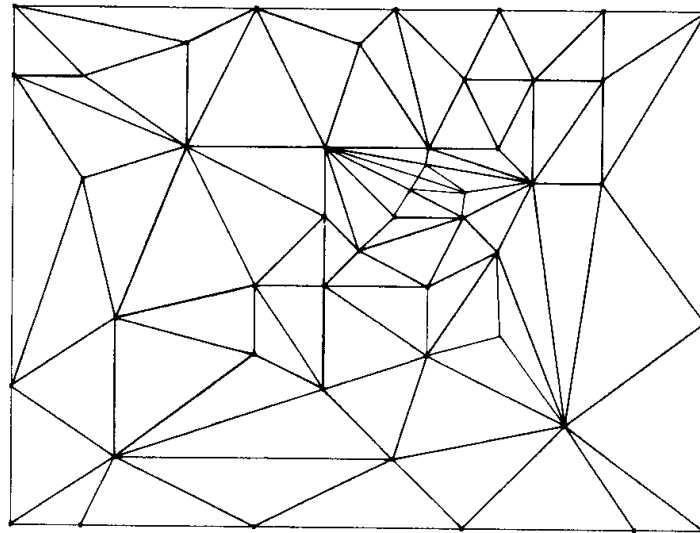


圖 3-2 基於 Delaunay Triangulation 的 DEM

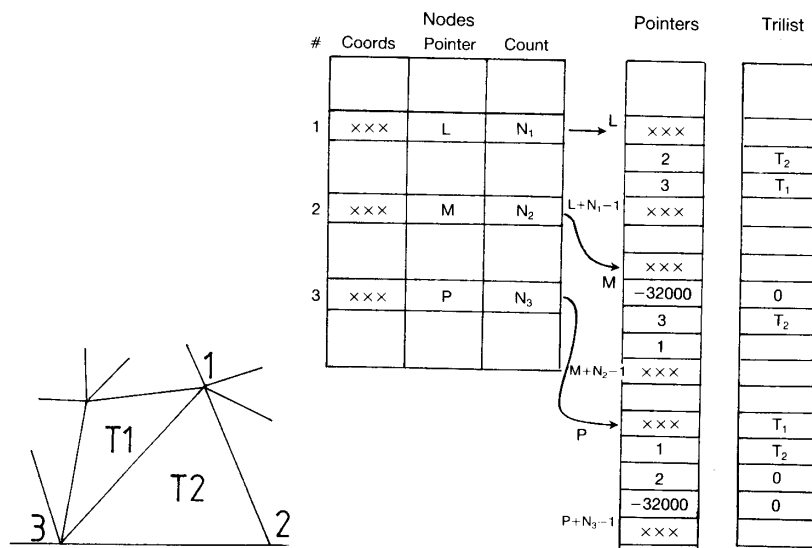


圖 3-3 TIN 的資料結構

#### 4. DEM 的資料來源及取樣方法(Data source and sampling methods for DEMs)

DEMs 所需的地理光學採樣方法：航照、雷達掃描

Selective sampling：在採樣過程中或之前選擇採樣點

適應取樣(Adaptive sampling)：在採樣過程中或於地面刪除多餘的點

逐漸取樣(Progressive sampling)：採樣及分析資料一起，由資料分析的結果指揮取樣進行

取樣：人工—太慢，容易有誤差

半自動—可增加速度及正確性

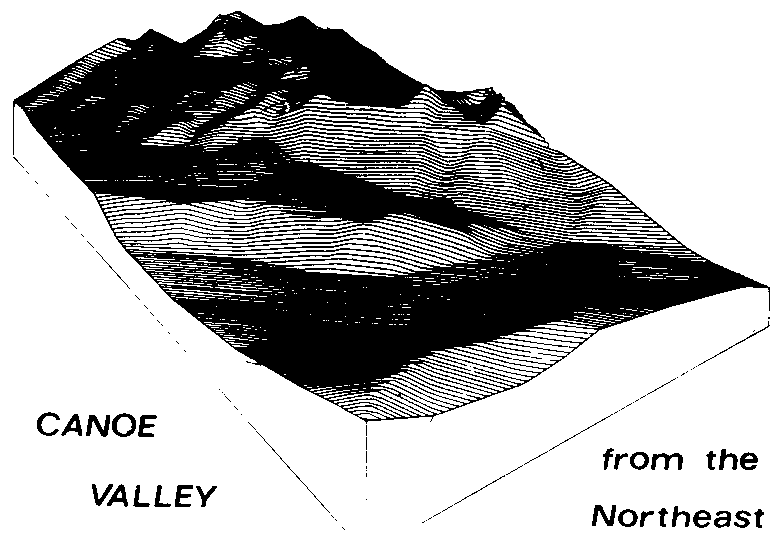
全自動—雖然很快，但不夠正確

採樣以不同方式進行端賴所需結果而定。Purposive sampling 將等高線 form lines 剖面及形態線數化。



CANOE VALLEY

(a)



CANOE  
VALLEY

from the  
Northeast

(b)

圖 3-4 從 TIN 模式結合成的 Canoe Valley 的等高線圖及區塊圖

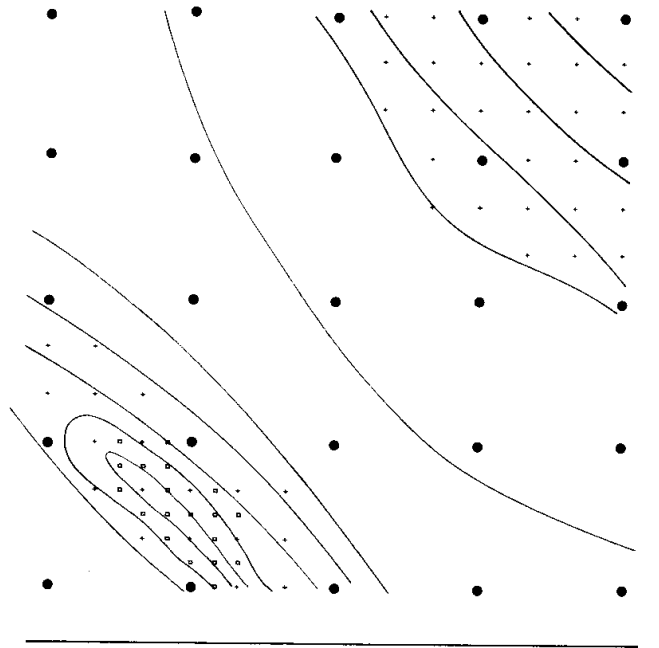


圖 3-5 在 Progressive Sampling 中，樣本點的密度可根據數值地形複雜程度自動調整

逐漸取樣(Progressive sampling)：自動隨地面複雜度的不同調整網格大小，一開始是較粗的網格然後增加其密度由電腦分析決定取樣點數目。本法通用於沒有不規則地區的情形。不規則的地區可用修正的逐漸取樣稱為混合取樣。

Composite sampling：首先，在採樣前，先手繪自然邊界或不規則物體之輪廓許多不連貫的改變無法用半自動逐漸或混合取樣達到這些資料就需用 selective sampling 最後以逐漸及混合取樣收集到的資料需轉換成第一的高程模式。

### Data registration and geocoding

數值化的立體航照因為高度變化、飛機傾斜造成變形扭曲，必須校正，最簡單的座標系是規則方格適於小面積大面積則常用 UTM。

### 5. 由 DEM 導出之產品(Products that can be derived from a DEM)

表 3-2 DEMs 導出產品

- 
1. 方塊圖：剖面圖、平地
  2. 數學積分算出體積
  3. 地形圖(等高線圖)
  4. 景觀視域圖(line of sight maps)
  5. 坡度圖、凹凸圖
  6. 地形圖
  7. 灌溉網路與灌溉基描繪廓
-

**Block diagrams, profile and horizons**

方塊圖(Block diagram)是 DTM 的最後歡迎形式之一，因其可將改變的值量化，方塊圖對於展示地貌設計或森林覆蓋模擬很有用。

**Volume estimation in cut-and-fill problems :**

DEM 可估計挖方及填方，由數值整合獲得體積

輪廓(Contour maps)：對於繪圖來說明仍屬粗略，以特別的數字法用來穿越高度矩陣的點連成等高線，如果太粗略以插入畫出較細的網格。

Line of sight maps :

在 landscape 上可看到點對軍事、規劃視域、聯絡網路娛樂上很重要

interrisibility maps : 由高程矩陣(altitude matrices) TIN

這個資訊可用於已有建物可看到地區的地圖及估算未來新建物可看到的區域(圖 3-11~13)

Maps of slope, 中凸(convexity), 中凹(concavity) and aspect(圖 3-14)

Gradient—高度變化的最大速率=slope

Aspect—改變的最大速率範圍方向=exposure

Convexity—slope 改變的速率=plan convexity=profile convexity

Concavity—負的 convexity

$$\text{gradient } \tan G = \sqrt{\left(\frac{\delta Z}{\delta X}\right)^2 + \left(\frac{\delta Z}{\delta Y}\right)^2}$$

$$\text{Aspect } \tan A = \frac{-\frac{\delta Z}{\delta Y}}{\frac{\delta Z}{\delta X}} (-\pi < A < \pi)$$

有限點 finite different 估算 gradient :

$$\text{最簡單(x 方向)} \left(\frac{\delta Z}{\delta X}\right)_{ij} = \frac{Z_{i+1,j} - Z_{i-1,j}}{2\delta X} \text{ 缺點: 局部誤差}$$

較佳估算式

$$\text{東西向} = \left(\frac{\delta Z}{\delta X}\right)_{ij} = \frac{(Z_{i+1,j+1} + 2Z_{i+1,j} + Z_{i+1,j-1}) - (Z_{i-1,j+1} + 2Z_{i-1,j} + Z_{i-1,j-1})}{8\delta X}$$

$$\text{南北向} = \left(\frac{\delta Z}{\delta Y}\right)_{ij} = \frac{(Z_{i+1,j+1} + 2Z_{i,j+1} + Z_{i-1,j+1}) - (Z_{i+1,j-1} + 2Z_{i,j-1} + Z_{i-1,j-1})}{8\delta Y}$$

mapping slopes : aspect maps 分 9 級，N, NE, E, SE, S, SW, W, NW 及平原地(flat gradient 變化極大 6 級平均 $\pm 0.6SD$  平均值 $\pm 1.2SD$  通常就夠了。

Image 會比原始表面雜訊多，可以雷射印表機畫去陰影或彩色改善或在畫圖前用局部移動平均將其平滑化。

Shaded relief maps :

人工的 Shaded relief maps : 利用明暗顯示三度空間的物體、人工慢、貴，需要高

度技術自動化：原理：由一個位置照亮一個模型，結果很像航照，但並不一樣。

1. Shaded relief maps 並不顯示 terrain cover，只顯示數化地表
2. 光源選在西北方  $45^\circ$  水平向上，這個角度易於理解
3. The terrain model 通常是平滑的及一般化的，並不保航照圖那麼詳細

Shaded relief maps 可以很簡單的產生，光亮度受光源方法、物質影響

Shaded relief 的計算不複雜，也不需要很大的記憶體除了展示之外

Shaded relief maps 之應用：

對展示單一影像的 terrain 特別有用，與主題圖的結合更能增進理解由衛星影像或網格化主題圖與 shaded relief maps 疊合的技術可應用在 GIS 套裝軟體

由 DEM 自動化地形描繪(Automated land from delineation from DEMs)：

當排水及橋線無法數值化分開時，可能需要用自動的過程由高度矩陣導出它們用高度矩陣發現線性特徵最大的問題是在數化的表面會由 noise 造成的陷阱，這個 noise 由於數化土地表面小範圍變異或原始資料量化的結果而產生。在估算坡度上，pits 可以用  $3 \times 3$  cells 的核心發現，一個 cell 定義為“pit”即如果它的高度  $\leq$  它的全部 8 個鄰居。

Detecting ridges and stream lines：

因為要偵測 ridges 必須定出全部  $\cap$  位置，河流線則反之。

定 catchment 的邊界：

本例中，只有單一的 catchment 邊界需被發現，以便由其他資料中分離出。

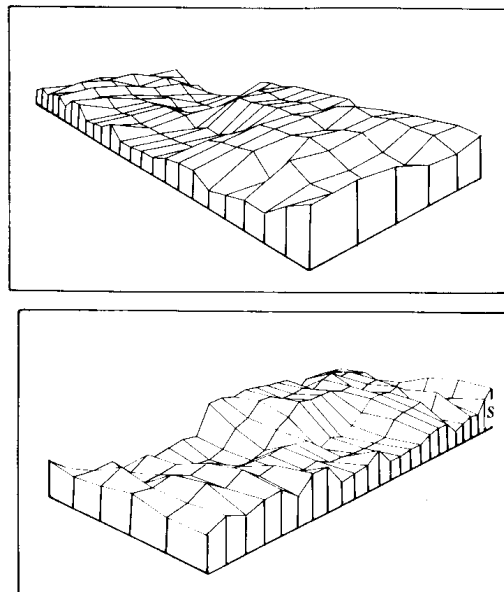


圖 3-6 一組在弗羅里達州 125 公尺網格中的  $6 \times 25$  的土壤剖面資料的第一主成分分數區塊圖

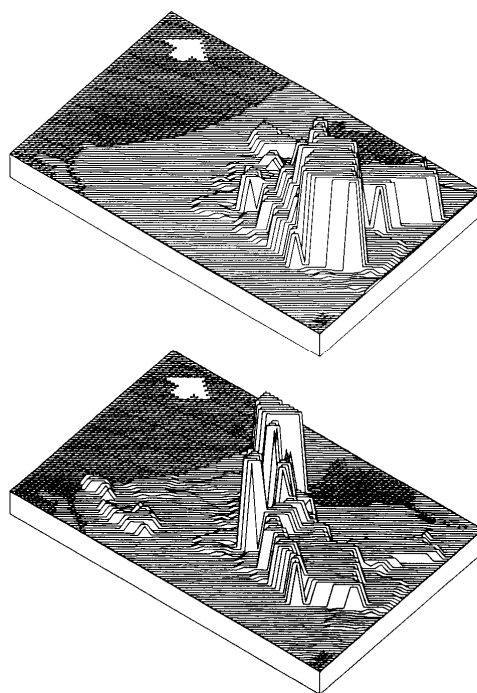


圖 3-7 來自主題圖的資料可以用區塊圖表示

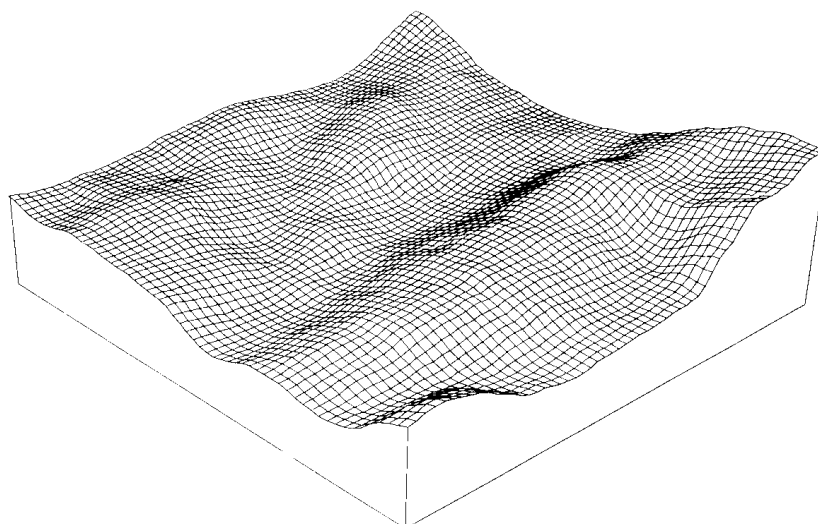
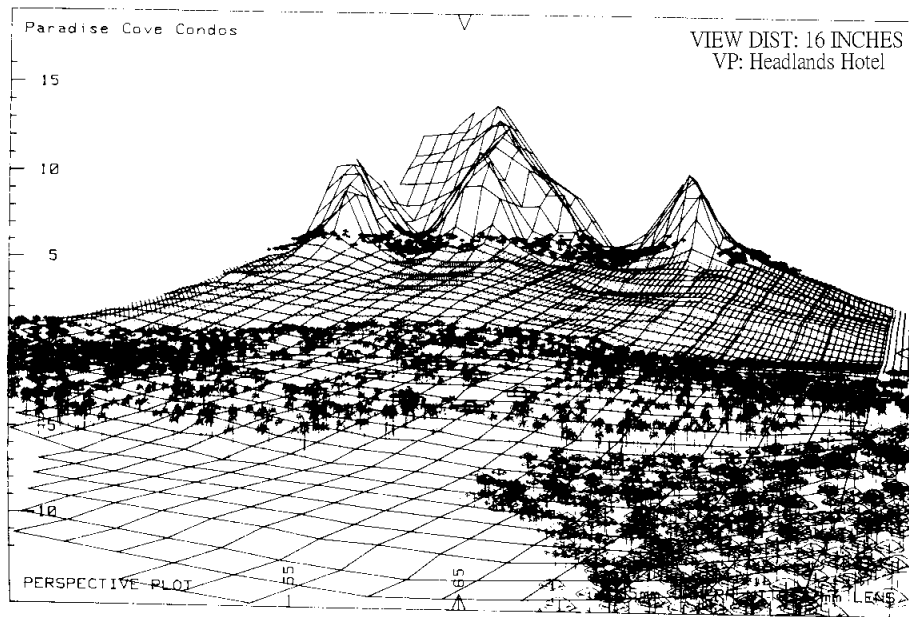
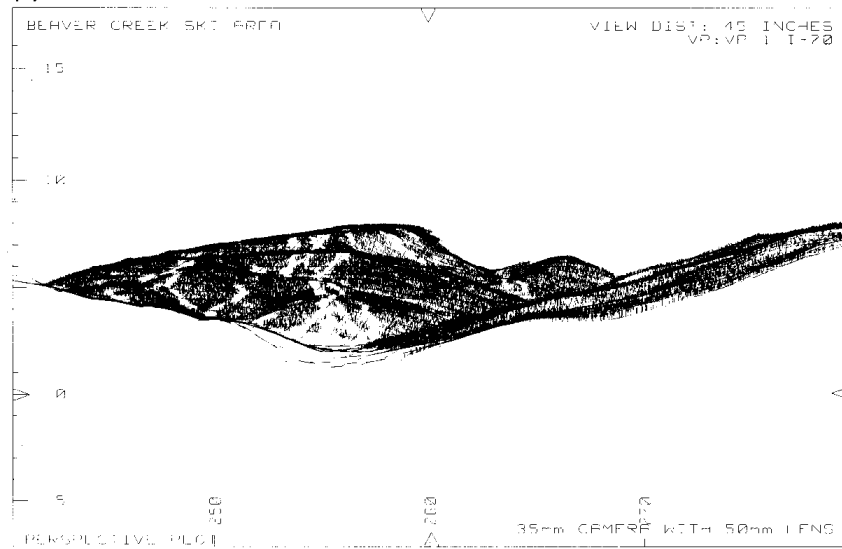


圖 3-8 Kissii District 地區 3750x3750 公尺中的高度矩陣經 60x60 點 DEM 計算的區塊圖

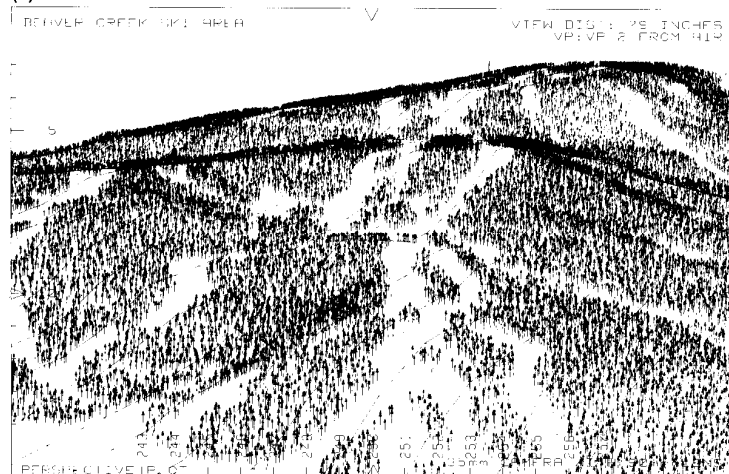
(a)



(b)



(c)



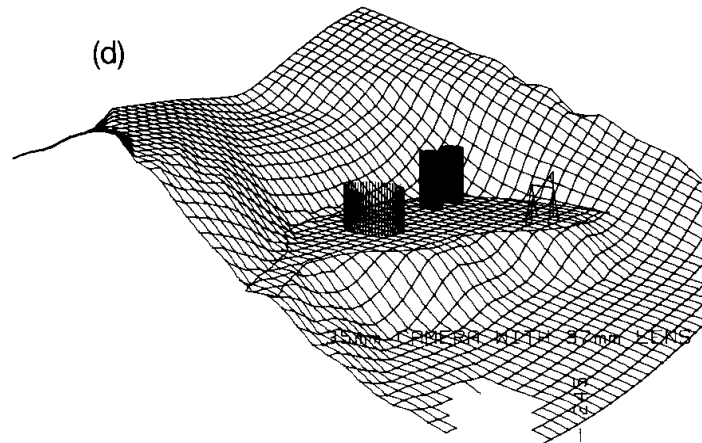


圖 3-9 區塊圖應用於表現地景細節：(a)模擬包括地景、熱帶闊葉樹等  
(b)、(c)概觀(d)地景中的石油和天然氣結構模擬

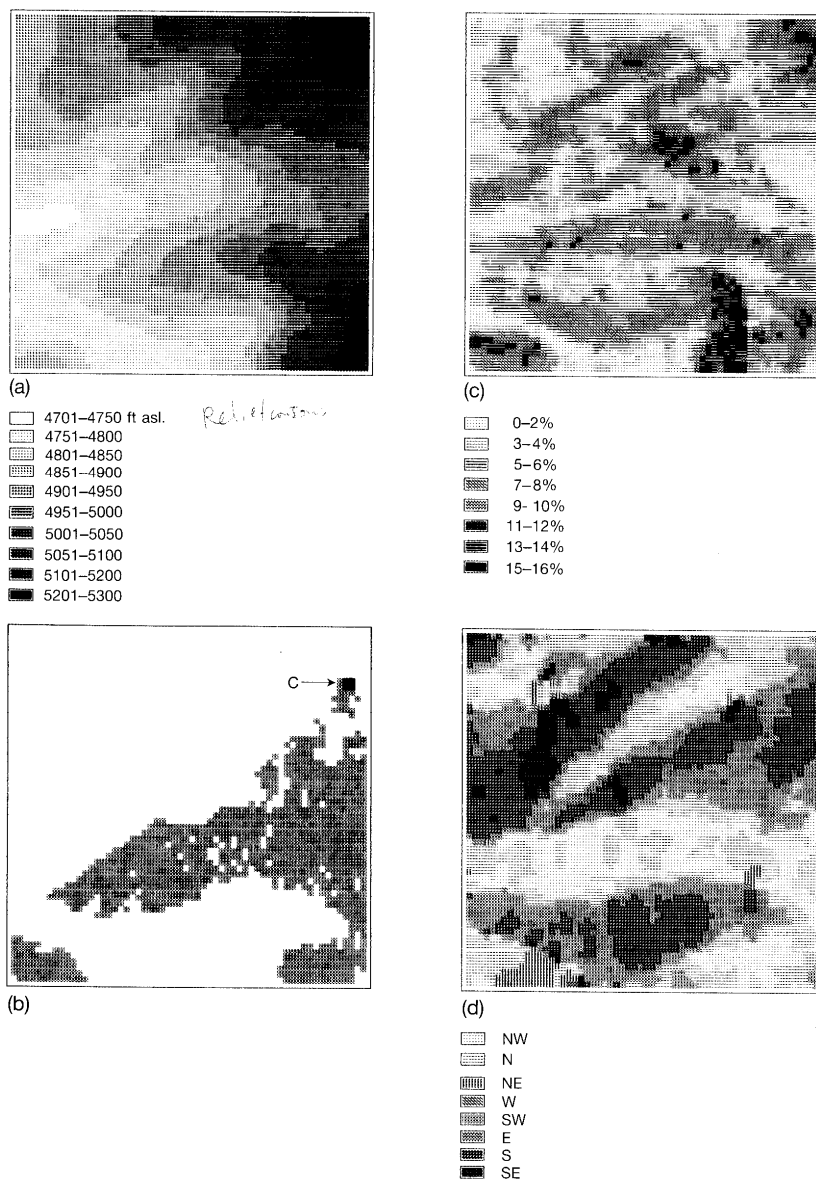


圖 3-10 Kisii 地區的高度矩陣以簡單網格地圖表示(a)等高線分佈  
(b)c 點咖啡工廠地點(c)斜度等級(d)方位分級



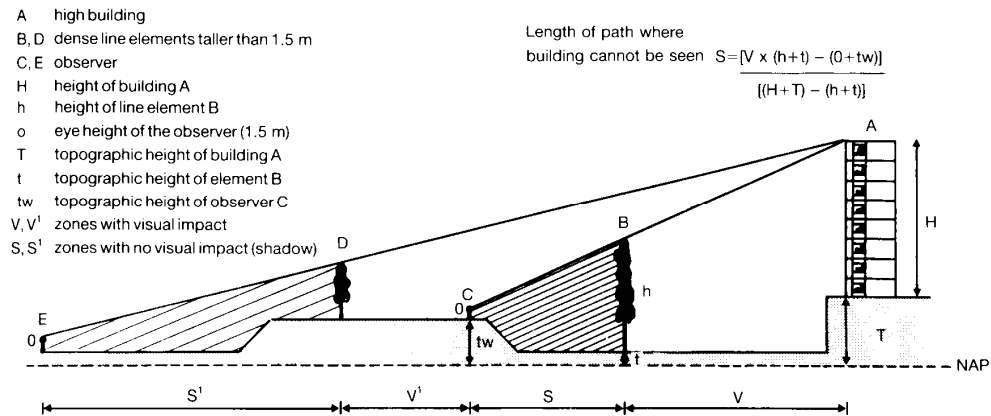


圖 3-11 荷蘭地景中高層建築可見度估計

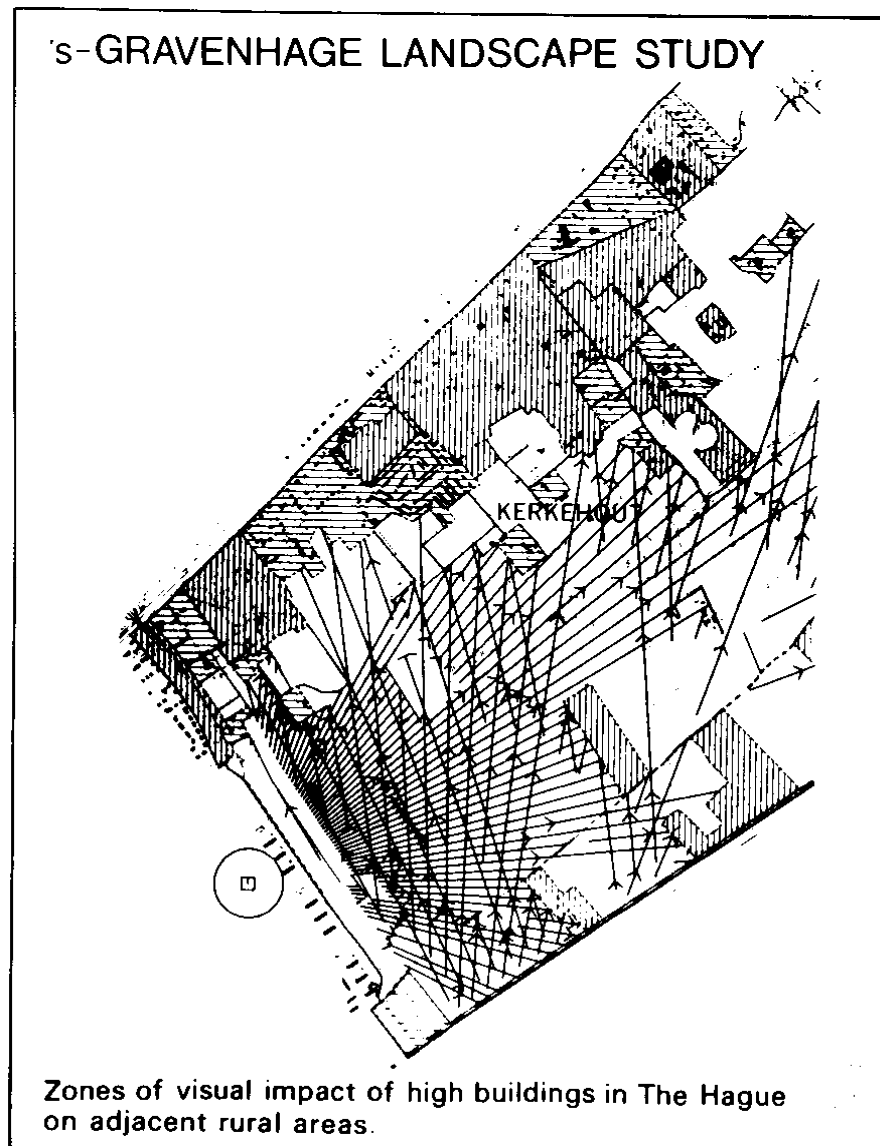


圖 3-12 道路的視域

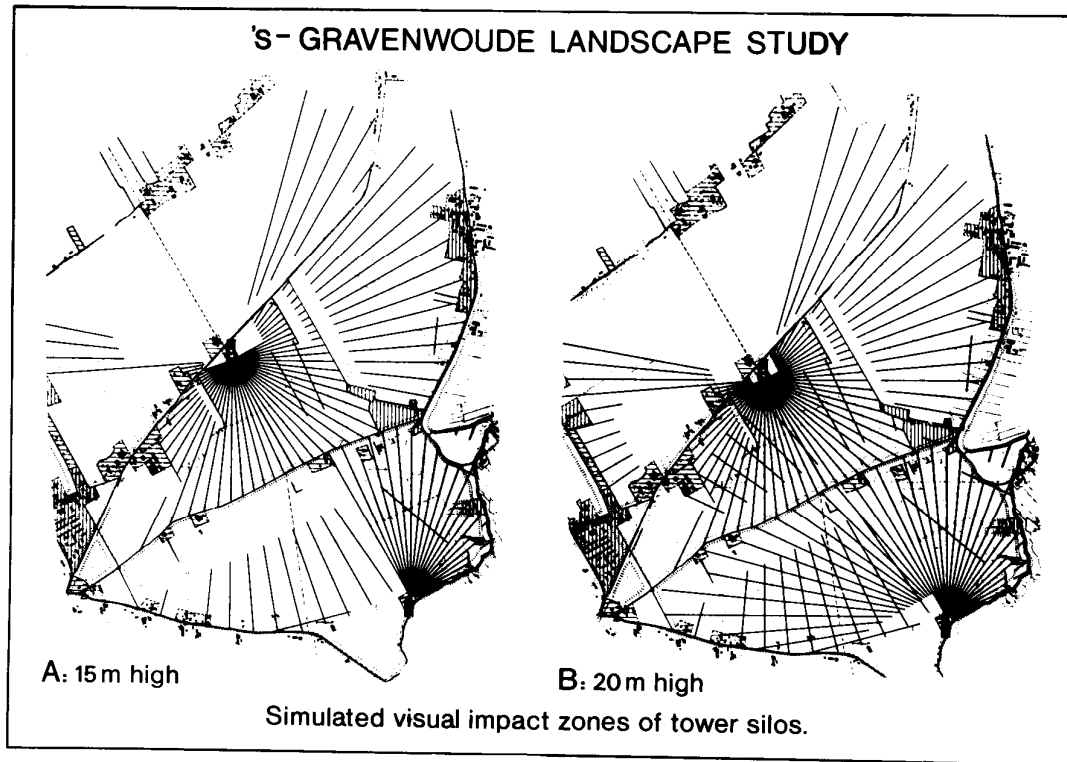


圖 3-13 不同高度之視域

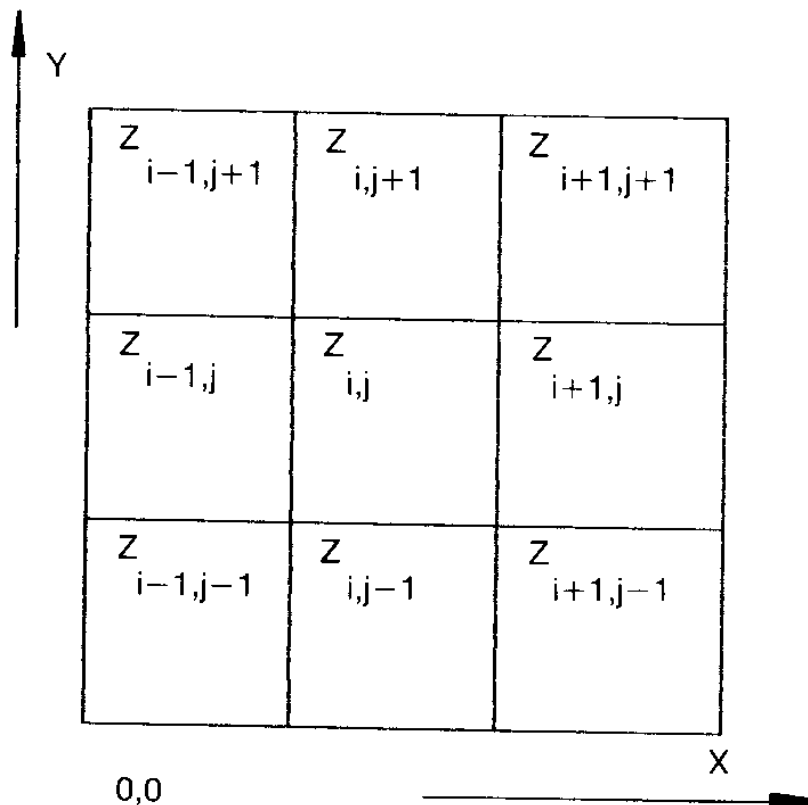


圖 3-14 Kernel 應用於計算高度矩陣的變化

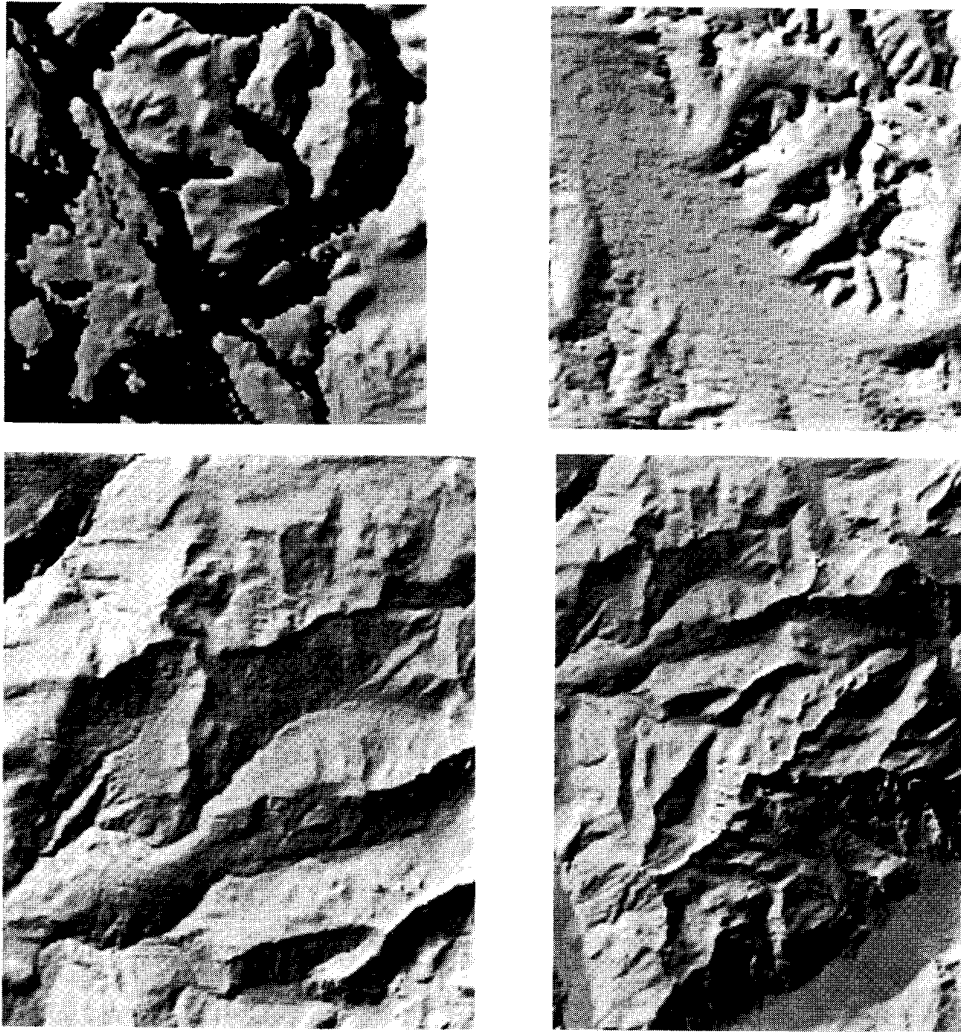


圖 3-15 使用高度矩陣的高品質圖片印表機所製出的地形陰影圖

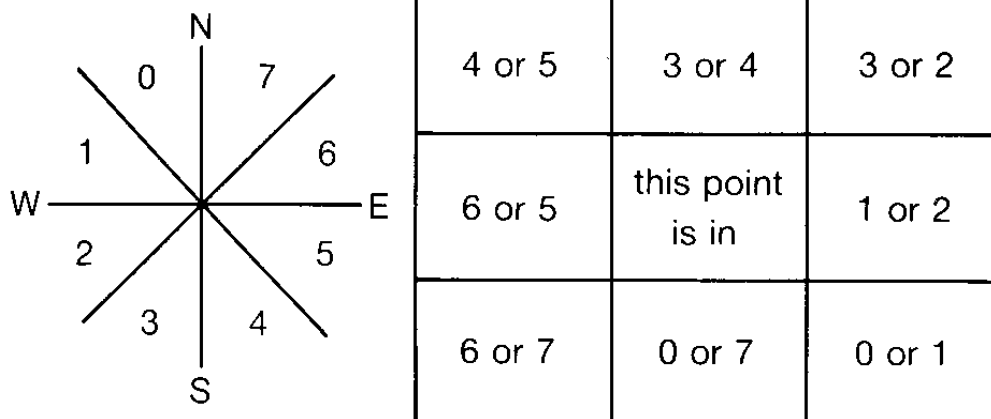


圖 3-16 方位分級及點相關位置的估計

## 第四章 資料的輸入、驗證、儲存與輸出

### (一)資料的輸入

資料的輸入是一種將資料編碼（encoding）及將資料寫入資料庫的過程。一個清楚的數化資料庫在 GIS 的使用上有極重要的地位，其亦為一項複雜的任務。GIS 的資料可分為兩個概念來說：（一）具有地理或製圖特徵的位置或地理資料需予以定義；（二）與該地理或圖徵的相關的屬性記錄。我們可利用這些地理或圖徵作為空間及非空間屬性資料連結的項目。

GIS 的資料輸入可以以下列三個主題來敘述：

- A.輸入空間資料（即數化）。
- B.輸入非空間資料，關連屬性。
- C.連結空間資料與非空間屬性。

其中每個步驟都必須經過資料確認及核對過程以確保最終所建立的資料庫盡可能減少錯誤。

#### A.輸入空間資料

並非只有單一方法可以輸入空間資料到 GIS 中。相反的，有許多種方法，只要可以相容的方法都可以單獨的使用或合併著使用。輸入方法的使用受到將來的應用、預算及欲輸入資料的種類所左右。而資料的種類即為現有的地圖，包括野外工作表、手繪文件、航照圖、衛星遙測資料、取樣點資料及由普查或其他調查所獲得的資料。這些空間資料多為不明顯易見的。

輸入的方法亦受到 GIS 的資料庫結構所影響。雖然在一個理想的系統中，使用者不必去擔心資料以網格或向量格式儲存與處理，但這畢竟與事實有一段距離，尤其是在有限的預算限制時。所以，接下來將討論不同組合的資料輸入方法及資料庫該為網格或向量格式。

#### 1.人工輸入至向量系統

我們所面臨的資料如點、線或面。這些資料的座標可參考已在地圖上的方格或經由方格圖或疊合方格而獲得。然後只需存入一個檔案或輸入程式中。

#### 2.人工輸入至網格系統

在網格系統中，所有的點、線和面是由一組小區塊組成，是最簡單且最常使用的輸入方法。首先先選擇一個網格單元的大小，接著使用一片具有選定網格單元大小的透明網格套疊到地圖上，並將網格內的地理屬性輸入電腦中。如圖 4.1。以文字檔的形式存入檔案，如 P58 左下。

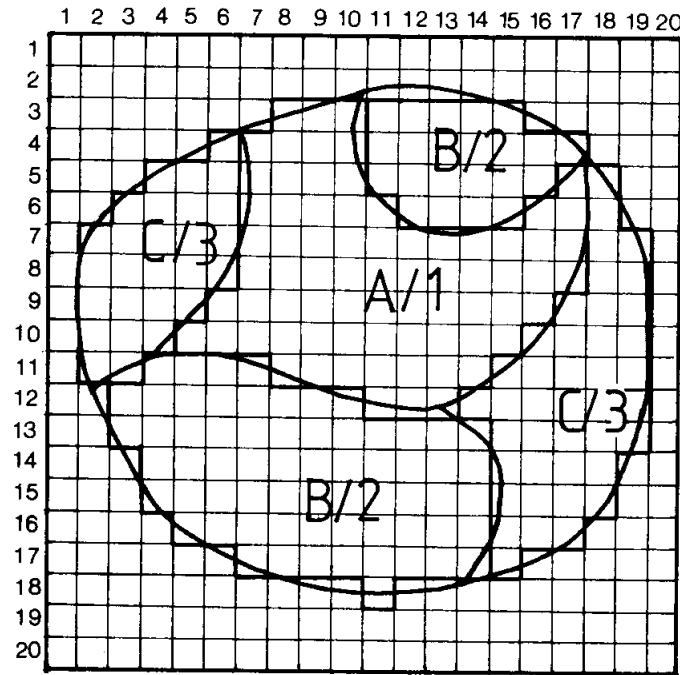


圖 4.1 以網格疊圖數化的一張簡單的地圖

A、B、C 為原始地圖編碼，1、2、3 為代表 A、B、C 的編碼

另一種記錄編碼的方式可由 P59 左上的種類長度編碼（Run-length codes），分別記錄行列出現的編碼值。這樣的編碼方式比起文字檔的編碼記錄方式少了許多的記錄次數。

### 3. 數化

經由數化可將我們所需要的點、線及面或網格之龐大數量的 X，Y 座標編碼。數化儀是一種電磁化的裝置，由一數化板及滑鼠組成。圖 4.2 說明數化儀的種類及外觀。其規格有從 27cm\*27cm 到 1m\*1.5m，有些有背燈。滑鼠的按鈕至少有一個，甚至有 4，12，16 或更多。

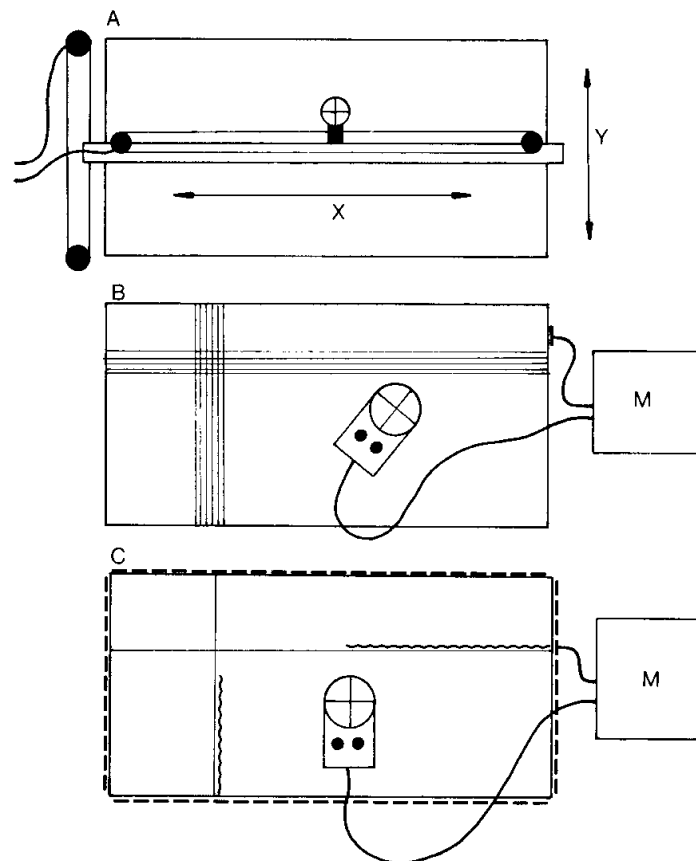


圖 4.2 不同種類的數化儀(a)金屬線數化編碼器(b)嵌入式金屬線數化儀(c)電波數化儀-微電腦控制

將地圖以膠帶固定於數化板上，以滑鼠定出四點邊界  $\text{Max}(X,Y)$ ，接者以滑鼠沿著地圖上的線條彎曲輸入  $X$ 、 $Y$  的座標。

向量轉換至網格 (V to R)

經由適當的應用程式，可將向量形式的資料轉換成網格形式，然而這其中將牽涉到資料遺失的問題，將會在第六章中加以討論。

數化的精確度受限於操作者的數化技術及數化儀本身的解析度。擁有  $0.025\text{mm}$  解析度的數化儀，其誤差不可超過  $\pm 0.07\sim 0.15\text{mm}$ 。

操作者往往求快而造成定址錯誤的現象。間斷的數化亦會影響數化的品質。線的結束點往往需要透過電腦的處理才能得到較好的成果圖。

#### 4. 自動化掃瞄

複雜且數量龐大的地圖以人工數化的方式必須花費相當多的時間，透過掃描器將地圖掃描成網格式資料。掃描器可分為兩種類型：網格式及直接沿線掃瞄。

網格掃描器：如圖 4.3

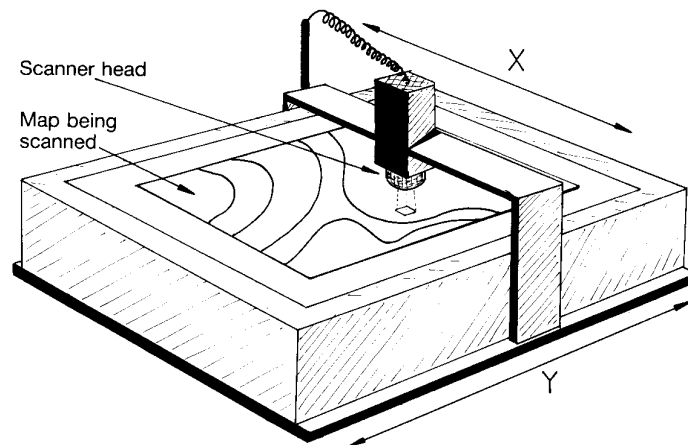


圖 4.3 平台式光學掃描器，掃描頭為 charge-coupled device (CCD)

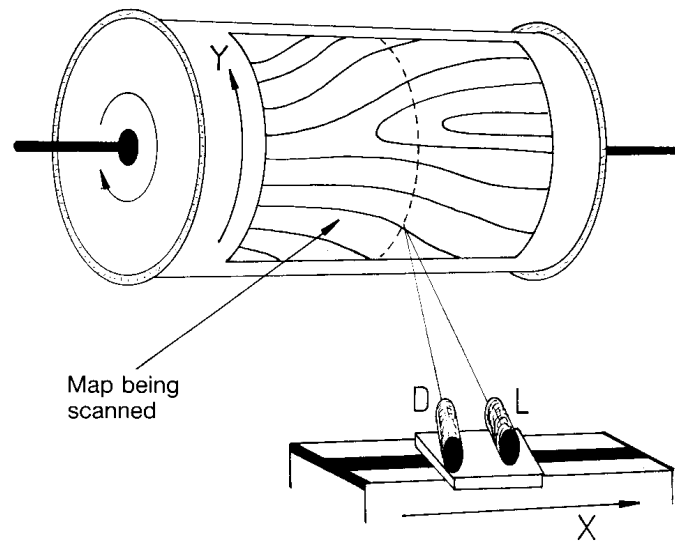


圖 4.4 滾筒式掃描器，Y 軸由滾筒決定，L 為雷射光，D 為偵測器

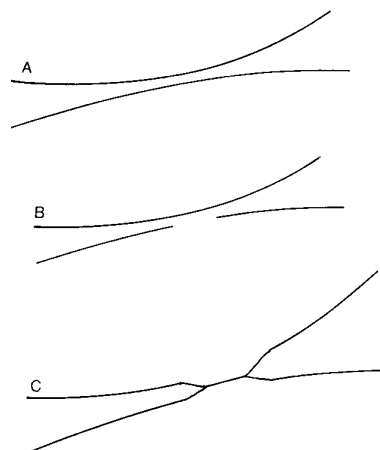


圖 4.5 等高線經掃描及處理過後可能發生的錯誤

CCD 是一種半導體，它可以計算落在其上的光子數量，其解析度可高達  $10 \times 10\text{mm}$  中有 250,000 個 pixels。

網格轉換至向量 (R to V)

網格轉換至向量是將網格的資料轉換成線的資料，並可經由筆型繪圖機出圖。如圖 4.6。

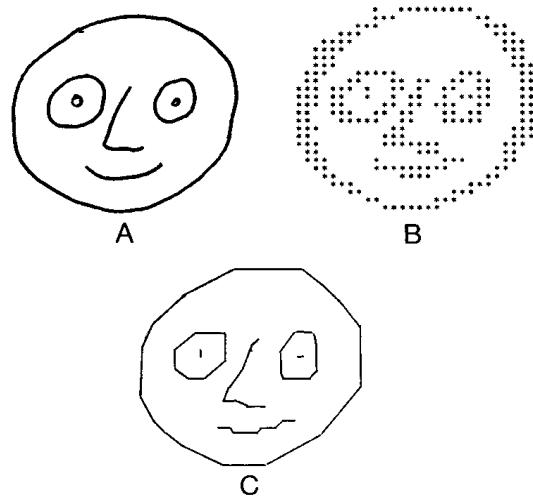


圖 4.6 向量化影像(a)原始影像(b)掃描後的網格影像(c)向量化及修正後的影像

可經由剔除的方式，將多餘的座標或線剔除，另外可看得見的線可經由 B-splines 的方法改善（見第八章）。

向量掃描器：雷射掃描，即雷射光束沿著地圖上的線掃描直到接和點或原來的點才停止。其優點為快速，並減少操作者操作的機會與時間，且在其掃描時就可以直接進行尺寸校正。

其他形式的地圖掃描器：錄影數化儀、立體解析繪圖機。

#### 5. 空間資料已存在於數位網格形式中

使用在低海拔調查的飛機上所裝置的衛星感測器或多光譜感測器將會產生地形的電子化資料，這些資料可在未被轉換成可見的影像地圖之前經由微波傳送到另一個接收站或可儲存於磁帶中。這些被掃描的網格中記錄著當時地球表面所反射出的光線波長。如 LANDSAT1 有四個波段（ $0.5\sim0.6\mu\text{m}$ ； $0.6\sim0.7\mu\text{m}$ ； $0.7\sim0.8\mu\text{m}$ ； $0.8\sim1.1\mu\text{m}$ ）如此便可分辨出植物、水或岩石的不同。

然而低解析度的 LANDSAT 影像需結合地形圖以提高其可信度，並可透過訓練樣本調整及處理網格的扭曲或變形問題。

6. 數位空間資料的其他來源：內差法所得資料、其他（具有空間資料的性質，但並必要進行空間資料的編碼）

B. 輸入非空間資料：非空間關連屬性描述這些空間資料的性質，如路的長度、寬度、路面質地等。對森林來說，即樹高、胸高直徑、林齡、樹種等。

C. 連結空間資料與非空間屬性：空間資料中的點、線及面都有獨立的辨別編碼，因此，可利用此一辨別編碼連結至非空間的屬性資料。



圖 4.7 為數化一組邊界及非空間屬性及其連結形成一地形資料庫所需的步驟。

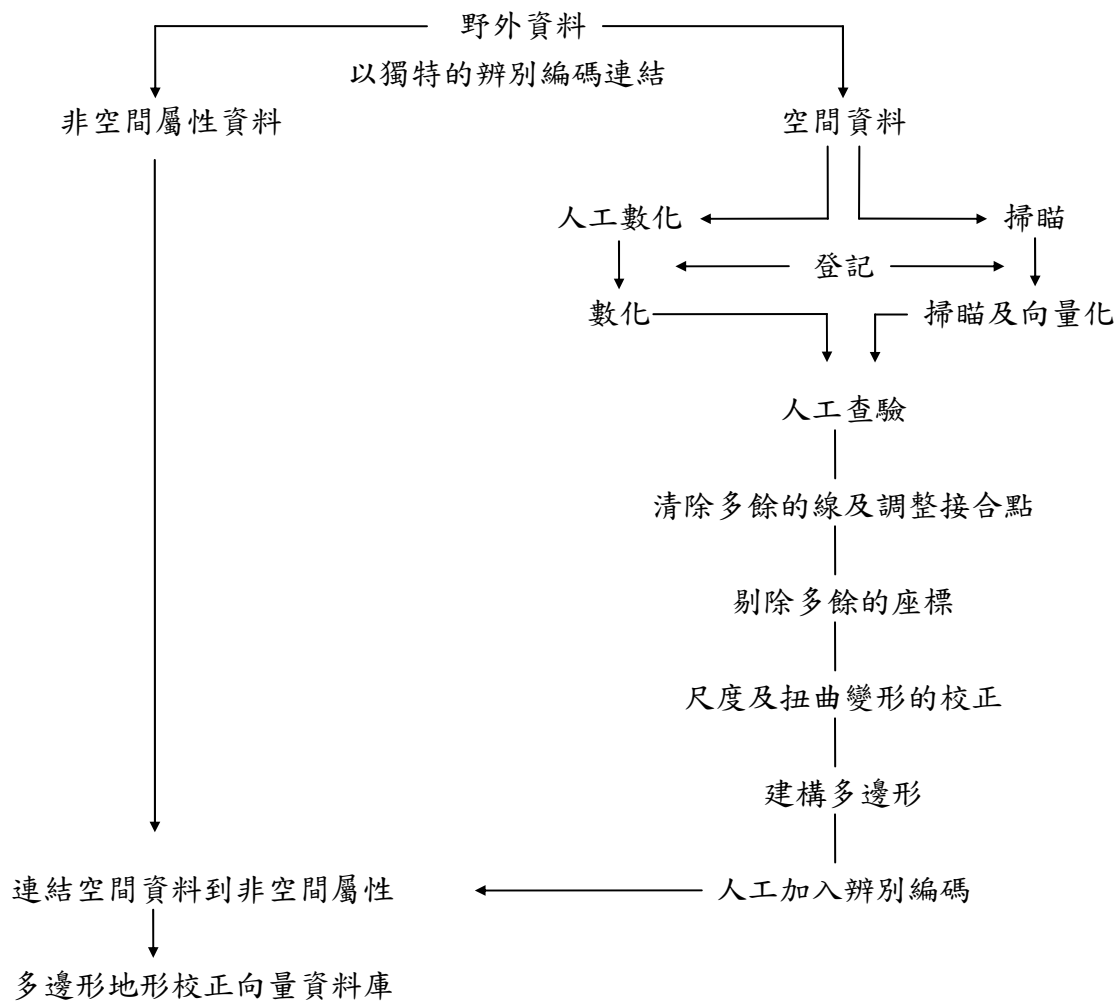


圖 4.7 建立地形校正向量多邊形資料庫的步驟

## (二)資料的驗證、更正與儲存

空間資料的編碼與非空間屬性資料輸入的過程中可能會發生錯誤，這些可能發生的錯誤如下：

- 1.空間資料不完全或重複：點、線或網格單元輸入時省略；網格轉換至向量時造成線段重疊；人工數化接合地方突出等。
- 2.空間資料位置錯誤：數化時不注意所造成。
- 3.空間資料的尺度錯誤：數化時便在錯誤的尺度下造成。
- 4.空間資料連結錯誤的非空間屬性。
- 5.空間資料扭曲變形：由於地圖紙張的折疊造成地圖的變形、資料轉換時產生的變形、飛機傾斜、地面凹凸起伏不同、焦距不同等。
- 6.非空間屬性資料不完全：輸入錯誤、野外編碼錯誤。

表 4.1 列出大部份 GIS 中資料的維護及處理以便於資料的驗證與更正。

表 4.1 一般 GIS 資料的維護與處理

加入/刪除/改變	交談式編輯接合點、長度、文字、字型、地物屬性 (點、線、面)
移動/替換	移動地物(點、線、多邊形或一組網格)到一新位置
縮放/矯正	調整座標以符合真實
轉換尺度	調整至一已知的尺度
縮放視窗	放大或縮減可見區域
剪下	將可見區域分成部份
合成/尖端接合	接合連續的線及文字訊息
多邊形疊圖及合併	圖層套疊
3-D 顯示	3-D 顯示資料
網格轉向量	轉換網格資料至一組線
向量轉網格	轉換線或多邊形至網格
一般化及平滑化	資料減量及資料結構轉換，剔除多餘的線及座標資料修補及公佈計算項目的簡單重複工作、報告面積及周長、簡單距離等、簡單布林搜尋、以文字檔記錄結果以工作更進一步的處理

#### 橡皮伸張法(彈性紙張)轉換及扭曲 (rubber sheet)

錯誤的地圖可想像為一彈性紙張可從各方向伸展，並與正確的基本圖上的點連結以得到正確的圖，如圖 4.8。橡皮片轉換不能用在網格資料，因為網格單元及資料是固定的。

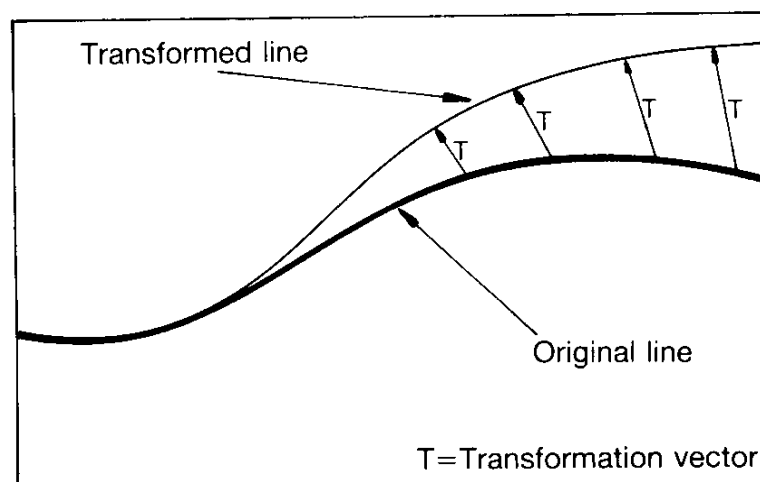


圖 4.8 橡皮片轉換的向量資料

網格資料的轉換可藉助簡單的數學內差法獲得解決。即假設分離的

方格資料為一連續的統計表面。因此，可利用內差法來求出新的座標軸；這就是影像處理中的「迴旋」。如圖 4.9 藉由周圍的四個網格計算出一個新值。

圖 4.10 利用周圍的 16 個網格資料及加權的方法求算出一新值。彩色圖 1~6 及 13、14 都是以這樣的方式求出。

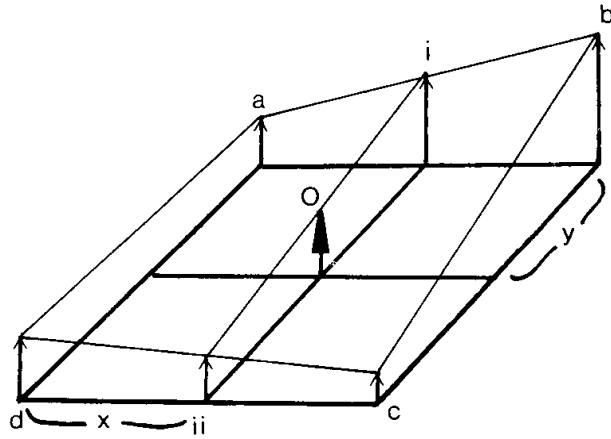


圖 4.9 雙軌內差法，以四個以輸入的點（a~d）求出新值（O）

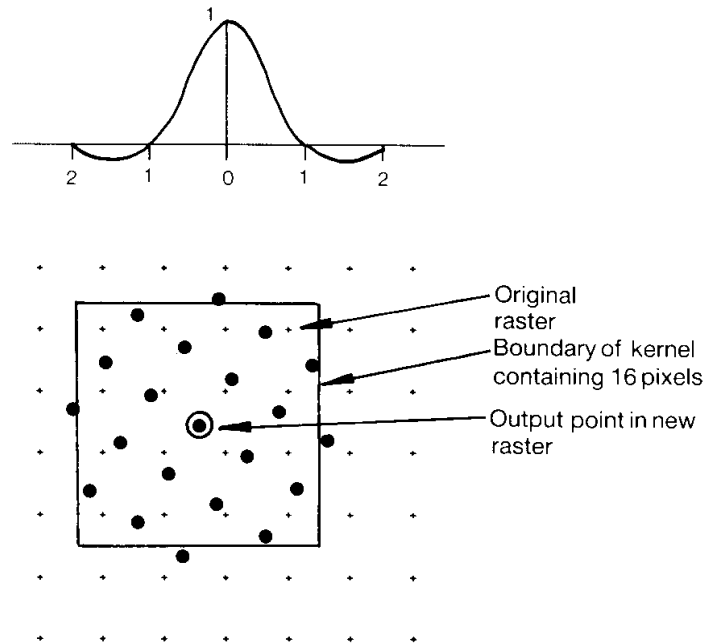


圖 4.10 利用正弦方程式的立體迴旋

#### 資料驗證

空間資料的驗證：以電腦半透明印出，並與原圖比對，找出漏失的資料、定位錯誤及其他錯誤。

非空間屬性資料：先以電腦掃描，並輸入超出的座標，讓電腦將大錯誤先剔除；另外亦可挑出辨別編碼的錯誤。耐心便可將大錯誤挑出，然較難的是語法上沒錯但不通順的問題。

### 資料編輯

對於錯誤的網格資料與非空間屬性資料可利用鍵盤及數化儀予以更正。其中系統的功能應包括移動、替代、刪除、插入、伸展及切斷。

網格式資料庫在多邊形網路中，如果線 X-Y 移動或改變，則多邊形面積需重新計算。

### 資料更新

有些空間資料並不是恆久不變的，如行政區域、田野邊界、景觀等。資料更新牽涉到重新調查及管理組織的問題。有些資料可以直接在原來資料庫更新，但是有些卻重新調查或輸入會較經濟。

### 資料備份與儲存

建立一數位資料庫花費許多金錢與時間，因此對於資料的備份與儲存非常重要。通常資料以磁帶的方式儲存。適當的磁帶儲存環境相當重要，磁帶種類與品質的不同亦有不同的保存期限。另外需要注意的是，不可有磁場在其附近。技術的進步，光碟片等高密度儲存設備有助於資料的備份與儲存。

## （三）資料的輸出

資料輸出是將所得的結果以一種使用者瞭解的形式呈現或允許資料傳遞到另一部電腦的過程。為人所習慣的輸出為地圖、圖及表的形式，而電腦所使用的確是以磁帶的方式，並可經由其他電腦讀取；或經由網路、電話線等傳輸資料的電子化資料。大部份的 GIS 包含表 4.3 的輸出資料形式的功能。GIS 的出圖方式常依靠電腦繪圖技術的發展。

表 4.3 一般 GIS 相容的輸出功能

選定地區的縮放
尺度選擇
顏色選擇
格式化文字、線、顏色及向度等
3D 顯示
顯示資料的圖層及套疊
繪圖指令
由磁帶叫出資料的指令
文字檔輸出的列印指令

為人所習慣的輸出裝置可將之分類為暫時性螢幕顯示及永久性出圖。暫時性螢幕顯示如終端機；永久性出圖如 hardcopies。但這兩種形式都可在被歸類為



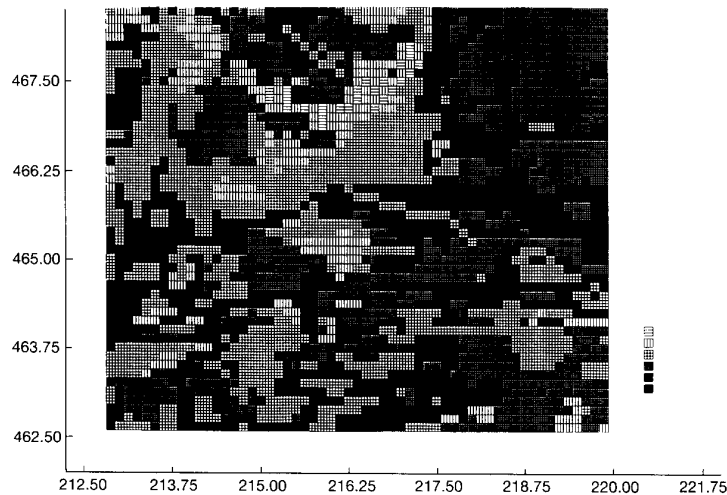


圖 4.13 網格結構資料經由筆型繪圖機出圖，通常需龐大的電腦配備

### 1. 向量形式出圖裝置

(a) 筆型繪圖機：筆型繪圖機是利用機器人製圖。繪圖筆沿著線的 X、Y 的座標繪出。筆型繪圖機的靈活性和速度受到所繪的字母及圖徵的複雜形式所左右。出圖的品質受到控制筆的馬達精密度所左右。品質最好的筆型繪圖機是經由雷射光筆直接寫在縮影膠片上。

(b) 向量視窗或儲存顯示：向量視窗與筆型繪圖機相似，主要的不同是筆由電子光束取代，紙以綠色螢光幕取代。影像寫入螢光幕會顯示到其被清除為止。

一個快速、半永久的紙張輸出影像可由 hard-copy 裝置獲得。hard-copy 裝置對於編輯地圖的部份或設計地圖的文字及圖徵非常有用。

### 2. 網格形式出圖裝置

(a) 網格繪圖機：早期常見的網格繪圖機是線條列表機或一台終端列表機。線條列表機最大缺點是其並非方格的單元尺寸（1/10in\*1/6in），造成 Y 軸的延伸。後來這個問題由於矩陣列表機的發明而解決。後來彩色列表機的出現及技術的改進已大大提昇出圖的品質及速度。

特殊膠片記錄器：彩色或黑白的網格資料可經由使用特殊膠片記錄器而得到高品質的輸出。與滾筒式掃描器的原理相似。出圖品質與高品質照片幾無分別。

(b) 網格顯示螢光幕：利用電視快速掃描投射在螢幕上的技術。而不同的地方在於其電子掃描光束的長度是受到一個電腦記憶體所控制，即 frame buffer。frame buffer 是一個很大的矩陣，裡面的每個 cell 都有 0 或 1 的值控制電子光束的開與關。矩陣的大小或所需記憶體可經由解析度推測出來。我們可以經由電子光束的快速掃描得到在螢幕上移動的圖片影像。較簡單的網格顯示螢光幕為單色，如綠色、黑色等。由網格螢幕所製造出來彩色影像的 hard copy 可利用特別的裝置繪圖而獲得。最好的結

果是使用軟片繪圖機，已可利用低成本的彩色噴墨列表機亦可得到一快速、可靠的結果。

交談式圖形工作平台：交談式圖形工作平台是一組控制電腦套圖或設計的系統。其乃是以交談式設計運作來作地圖編輯細節或檢驗出圖前的地圖，通常由鍵盤、顯示器、小型數化儀、及一個或兩個高解析度的網格或向量顯示器。兩個顯示器的目的在於一個可看到全圖，另一個顯示細節部份。雖然大部份的互動式圖形工作平台使用向量資料庫及 CAD/CAM 的系統介面，傳統的商業發展較偏好高解析度彩色網格顯示器來取代單色顯示器。有些系統結合網格/向量螢幕，向量螢幕顯示固定訊息而網格顯示彩色，使得使用者可以看出線條與圖徵是如何分佈。這樣的顯示方式在編輯時非常有用。

圖形輸出軟體（一軟體對一硬體→一軟體對多硬體）

直到最近，大部份的圖形輸出才都是裝置相依的，也就是說，可是之前加入一項新的輸出裝置，應用程式就要變更。由於國際需要，已朝向裝置獨立發展，如 GKS。

### 3. 資料展示

大部份的 GIS 使用者都希望由系統輸出的圖形與傳統的地圖產品相似，如平滑的等高線等。比例尺、色調的使用亦非常重要。電腦技術的進步及軟體的開發將有助於資料的展示。

### 4. 總結

輸出裝置可由很便宜到很昂貴。大致上說來，功能越強，出圖品質越好也就越貴。網格及向量展示設備的比較可由表 4.4 看出。

表 4.4 網格及向量展示的比較

裝置	型式	優點	缺點
繪圖機	噴墨	快速、便宜、品質低 無毛邊 體積小	列印過程穩定、只能使用最貴的紙張（銅板紙） 較好的軟體昂貴
	筆	顏色完備 耐用 安裝容易	陰影部份列印慢，與噴墨比較慢 機器部份有時會影響精確度 高品質的出圖昂貴
展示	充電式	使用便宜的電視技術 如顏色、動作	螢幕顯示品質低
	儲存式	高記憶容量呈現高解析度	通常為單色，彩色較昂貴

使用者介面：朝向人性化使用者介面的設計，大容量且便宜的儲存設備有助於這個目的的達成。另外不同的應用程式有不同的使用者介面。

## 第五章 資料分析及空間模式建立的方法

### 1.前言：

GIS 與電腦輔助製圖的主要差別就在轉化空間資料以便於解答特殊問題的能力。圖 5-1 顯示在 GIS 資料利用及分析能力的主要種類階層總觀。

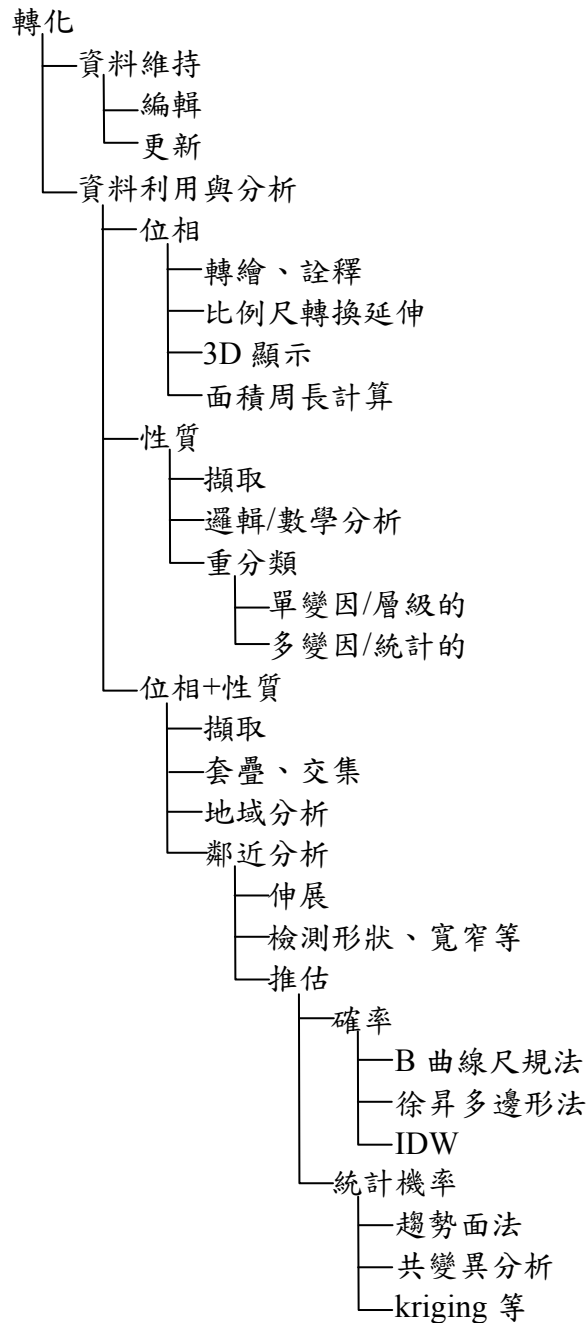


圖 5-1 地理資訊系統中資料轉換作業的層級

資料分析的一般問題如圖 5-2，使用者有問題，GIS 資料庫來解答問題介於資料庫與輸出間建立一個連結(link)，本章即在說明連結如何定義及使用。



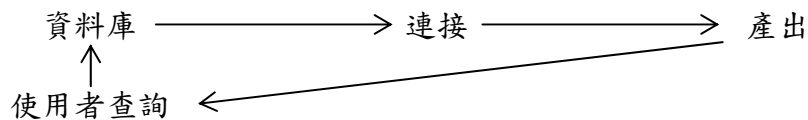


圖 5-2 資料分析之程序

## 2. 資料庫的定義：

資料庫可提供所需資訊資料，在資料庫中存在形式是點、線、面及其屬性。

## 3. 簡單資料擷取：

如果資料利用覆蓋或層狀結構被編碼(encoded)在向量系統內，在圖層在一起的資料就可很容易獲取。

布林邏輯(Boolean logic)：

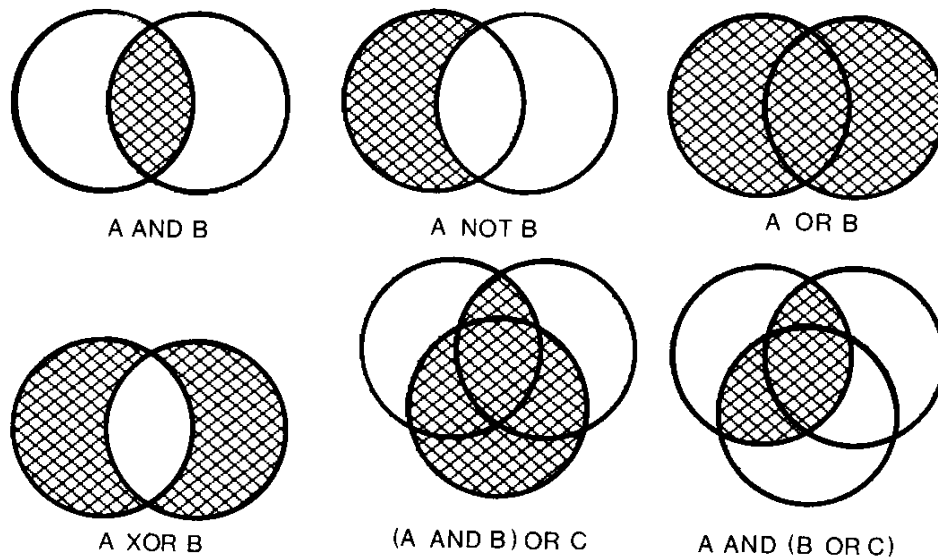


圖 5-3 應用布林邏輯去擷取

### 3-1. 重分類及展示(Reclassification and display)：

地理資料尋取後，通常需要重新分類及重新展示，並確認鄰接的多邊形有相同的名字，他們之間的界線取消，新的多邊形網路建立。

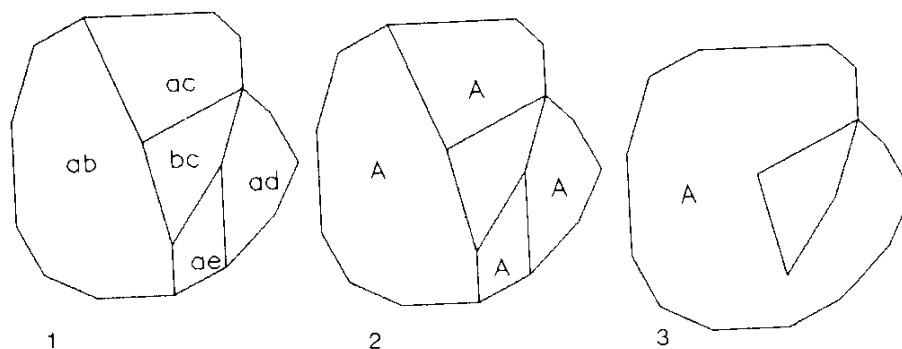


圖 5-4 重分類(1)來源(2)重分類(3)移除不需要邊界

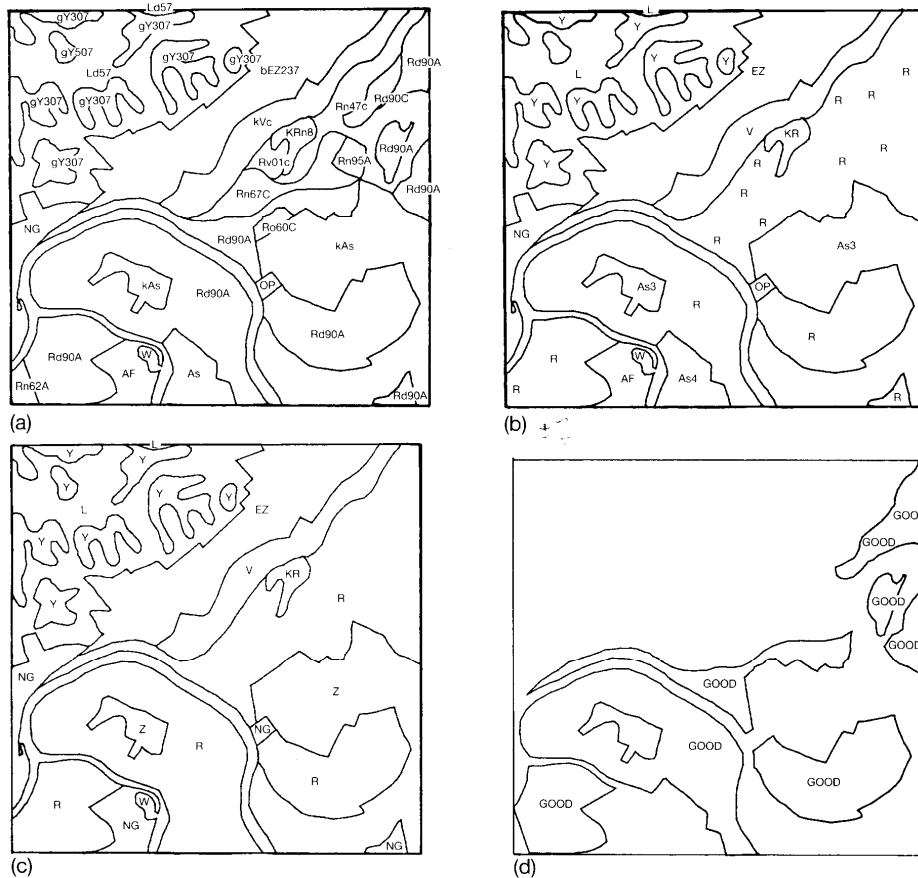


圖 5-5 土壤序列圖(a)荷蘭土壤測量偏 左至右表示小寫：表土，大寫：土系，  
小寫：排水情況，數字：結構級，As 為複合組成  
(b)土系(c)b 圖之合成(d)布林運算自土系"R"排水"d"面積  
>12ha 為"GOOD"

### 3-2.Boolean operations on two more maps :

GIS 有能力把 2 個多邊形網路貫穿，boolean logic 可應用至此結果。最後地圖多邊形數不只跟原圖多邊形數有關，還跟邊界形式有關，原圖邊界越旋繞；獲得的多邊形越多。

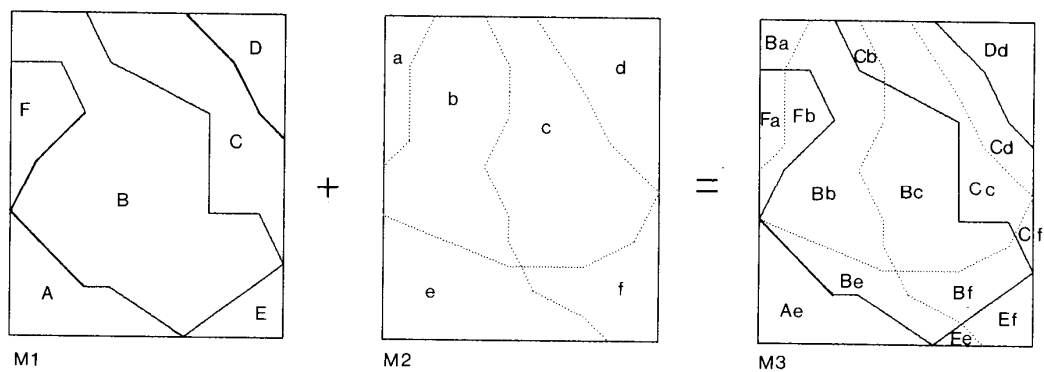


圖 5-6 兩個簡單地圖疊合成第三個許多小且新的多邊形

問題 1：電腦繪圖上如何重疊 2 個以上的多邊形網路？如何處理大量小而不

重要的多邊形，過與不及會造成資訊太少或產生不必要的資訊。

問題 2：向量形式多邊形疊合時，二地圖邊界線尋找相交時計算需求很重，重新連線及建立新的多邊形網路與其屬性一起。

#### 4.圖層套疊的一般方法(General Approach to Map Overlay)：

A region：同一屬性的一群 loci

The loci：一群 X, Y 座標

最簡單的 region 是一個點，只有一個 X, Y 座標對，最複雜的包括許多 loci。

地域也許包含了一個以上的多邊形，可能中間被穿孔或其他形狀(圖 2-11 d,e)

屬性：地理變化或性質可以二進位、十進位、區間或比例範圍表示。

An overlay：數個獨立之鄰近區域以特有屬性連接，不包含鄰接 region 相連在同一區域，每個 overlay 由一屬性所定義，與 chap2 描述的空間資料網格類似，唯一差別就是網格資料結構空間，必須量化成正方形 cell。

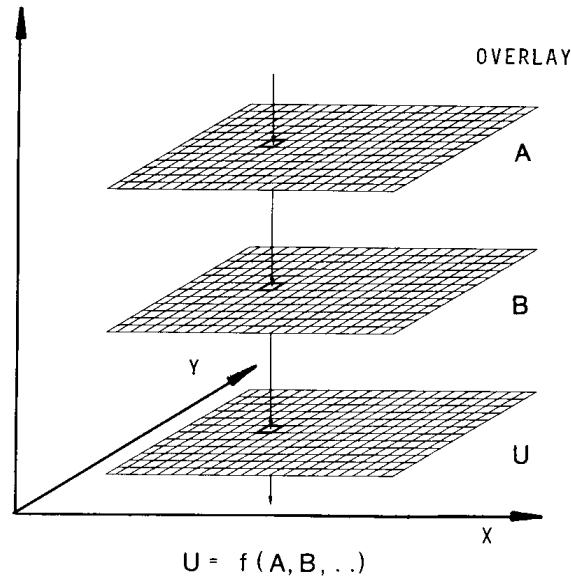


圖 5-7 套疊及變質

在圖 5-7 中，每個 locus 新屬性產生如同重疊屬性的函數  $U=f(A, B, C\dots)$

Classes of transformation function：paints and neighborhoods：

三級用來控制  $U=f(A, B, C\dots)$  轉換函數

第一級：操作只在 locus 有關連屬性，與鄰近 loci 屬性值無關

最簡單點轉換是加、減、乘、除指數等數學運算。

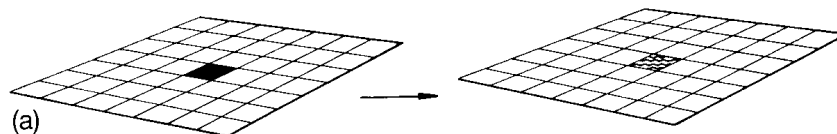


圖 5-8 (a)疊圖分析的轉換作業

第二級：與 region 的性質有關或一已知 locus 屬於的集合，這些性質與長度，面積、周長或形狀或與 loci(發生在已被 region on another 定義之地區有重疊上有某值)數目有關。

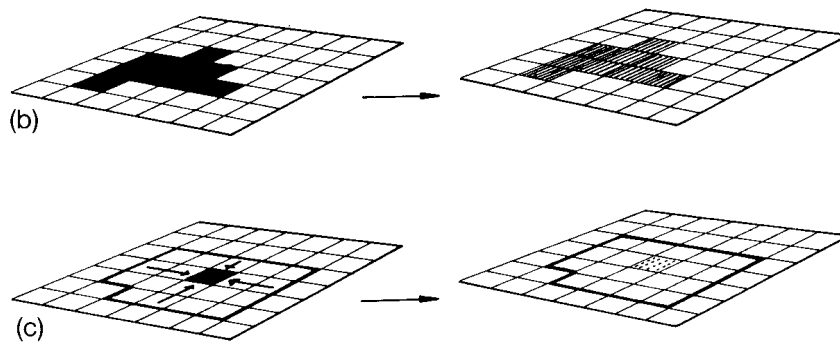


圖 5-8 (b)重新計算予不同顏色  
(c)由點值反應面的性質

第三級：點與某鄰居有關，包括了鄰居的點是否與問題中的點有相同的值或疊合後，問題中的點至不同值點之間的距離。鄰居函數包括了計算值，用加權平均、最大值、最小值、商、改變比例。

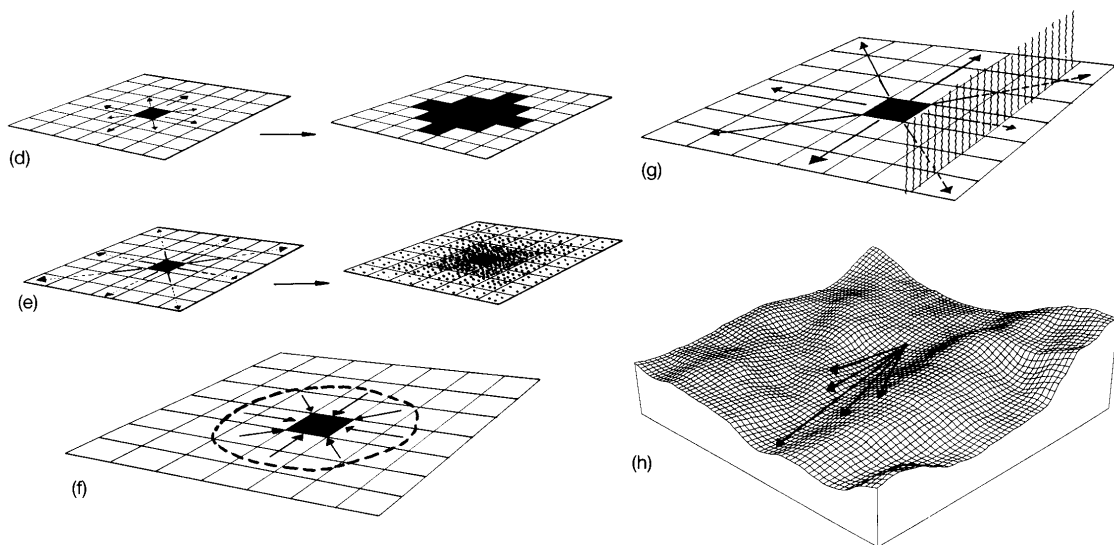


圖 5-8 (d)點間四邊做等值的擴展 (e)與距離的反比方式予權重  
(f)由周圍區域來推估中間的點 (g)由一點到其邊界  
(h)由一點到面

第四級：空間形態隨時間之變化。

表 5.2 套疊分析功能

(a)點運算

運算	關鍵字	修改號	摘要描述
求和	加法指令	權數	全部軌跡加法指令值的二或更加

減法		權數	(多)屬性給新的值
乘法	乘	權數	減
	EXPONENTIATE	權數	乘
		權數	除
乘			乘冪
三角函數			三角
對數函數			對數
平方根			平方根
平均	平均		整個取(加權的)平均的二或更加
			(多)值給新的值
掩蓋的	覆蓋		覆蓋一圖以非零單元值的第二的圖
			給第三圖
布林運算	交叉	布林	運算組合二或更加(多)分離式佈
			線電路或組合不斷的圖進入一個第
			三使用布林邏輯接通歸因於產生一
			個新的圖
發現極端事物	最大值/最小值		FIND 指令最大的/最小的值的歸
			因於於二或更加(多)覆蓋和寫給輸
			出的覆蓋
取出/隔離	抽出		選擇指定值和/或範圍的值自一覆
			蓋製作一個新的覆蓋
決定頻率分	直方圖		為覆蓋或選擇的區,計算直方圖
配			
匹配頻率分	分布延伸		為圖或部份的區,重算影像,因此;
佈至理論分			Z 值的分布匹配理論的分布
布			
指派一個常	常數		指派這不變的值到全部軌跡接通
數			
雙向的	單位刻劃	平均	交叉製表二 TOTAL 系統的屬性值
比較		H 總	自二圖接通一個點 BY 點基礎
		最大值	結果摘要的
		最小值	統計學產生一個新的圖
		中值	
		多數	
		少數	
		變化	
		偏差	
		重複	
多變元	主成分比較	最近鄰點	主成分分層輸出
	群	分級平均	族群分析
	分配	平行六面體法	
		最大概似法	以分類或分級標準將"單元"併至"
			級"

Fourier 變換+ 過濾	FOURIER		計算二維 Fourier 為貴賤零組件座 標變換域和應用濾波器 重計經由通用的 Fourier 變換圖藉
-------------------	---------	--	--

## (b)'Region'或說明單元(部件)運算

運算	關鍵字	修改號	摘要描述
重分級或重 新編號 決定 決定區或孔 依面積分區 套疊兩多邊 形網 區域內布林 搜尋 依包含點分 區	重新編號  記錄大小、面 積、周長、長 度、個數  Euler SIZE	線面	指派新的值指定單元( 部件 )或說 明值或範圍價格至圖 自資料庫讀取或計算域資料並顯示 或記錄  套疊兩多邊形產生新圖  從點  從點

## (c)鄰近運算功能

運算	關鍵字	修改號	摘要描述
連續性態值  找距離內位 置 連續面功能 點推連續面  等高線 推估  視域或 最短路徑 功能	CLUMP SCAN  SPAN DSEARCH SLOPE ASPECT PROFILE SPREAD  COLVTOUR TREND INVERSE OPTIMAL( 最 重) VIEW STREAM PATHWAY	距離判據 距離、平均、H 總、極大、極 小、中數、眾 數、最少值、變 異、機差、比例 距離 距離範圍方向  平均 最大 距離  覆蓋、向上、向 下、穿過  OVER OVER	為群或個體建單一極識成一新圖 依鄰近或某距離內統計之概述值產 生連續面圖  H 算像元之距離圖 找連續或不連續面內某距離或範圍 的點 坡度 坡向 剖面 從點或線，依其距離對各位置重計 其值 依距離或方向柵欄推估  從點資料估等值線 計算趨勢面 計算權重平均值 利用共變異方程協助推估  三度視域分析 找兩點間最短路徑 以路網時間、花費找最短路徑找點

點在多邊形	PPOINT		包含的面
-------	--------	--	------

### 5.Cartographic modelling using natural language commands :

人類如何簡單的與電腦溝通，自然的語言最好電腦能聽懂。

Map analysis package (MAP) : Dana Tomin 定義代數而寫了一個實驗電腦程式在 map 內，每個疊合描述一個分別的屬性，有單一的名字，每個空間分析操作一個動詞呼應，這個動作由一個適當的 modifier 控制。

### 6.Linking command sequences into cartographic models :

使用者面對問題時，要先提供答案所需的資料，接著使用合邏輯的過程把資料做出解答。

開發中國家地圖模式(cartographic modelling)應用在土地評估的三個例子：

例 1：土地品質(a)不限制(b)中度限制(c)嚴重限制

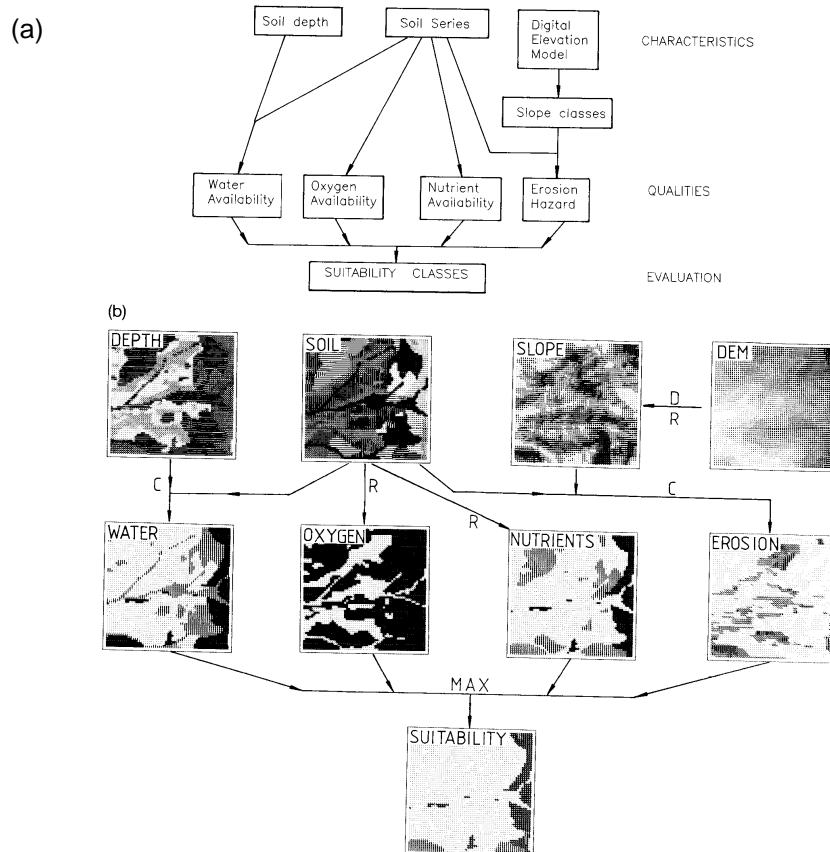
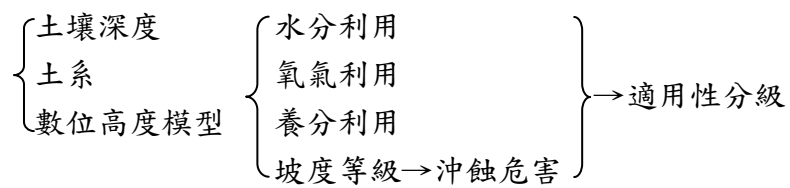


圖 5-9 FAO 土地評估流程圖

C : CROSS 經布林運算產生新圖

R : RENUMBER 重分級

D : SLOPE MAX : 最大值



RENUMBER – A→B

CROSS – A, D→B

MAXIMIZE – 所有屬性疊合取最大值→適用性分級

例 2：種玉蜀黍 40 年造成土壤沖蝕

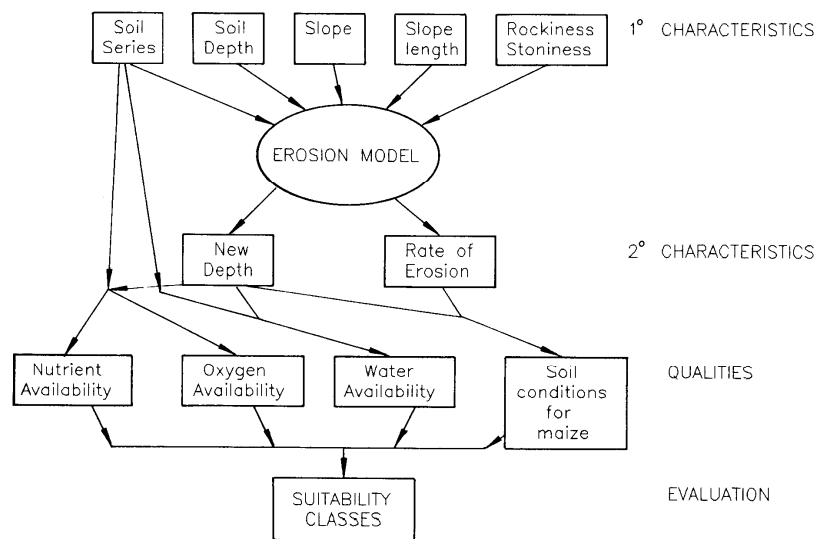


圖 5-10 估計地區適應性(對 40 年玉蜀黍經營)之步驟

估算土壤流失的 2 個經驗模式



MODEL:

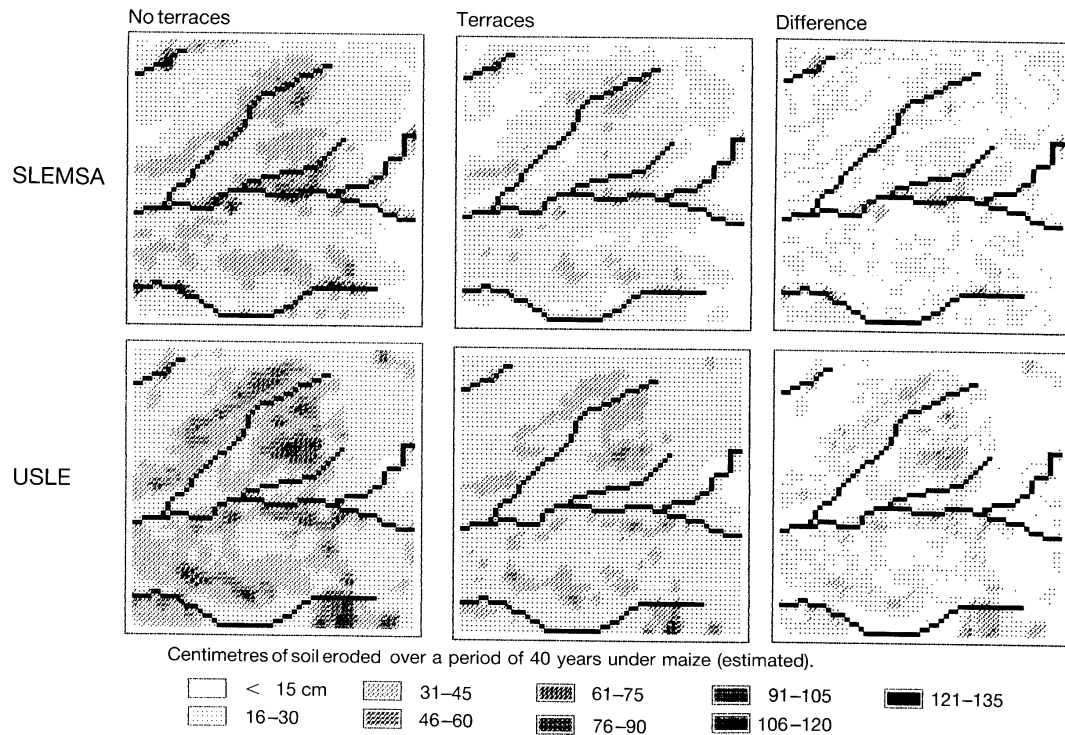


圖 5-11 40 年玉蜀黍經營土壤侵蝕公分數估計圖

USLE, SLEMSA 其缺點是太簡單、不夠精確

例 3：咖啡收益：

咖啡價錢(土壤圖→適合性)——運輸成本(道路、DEM(高度、坡度)→接近區)  
= 收益

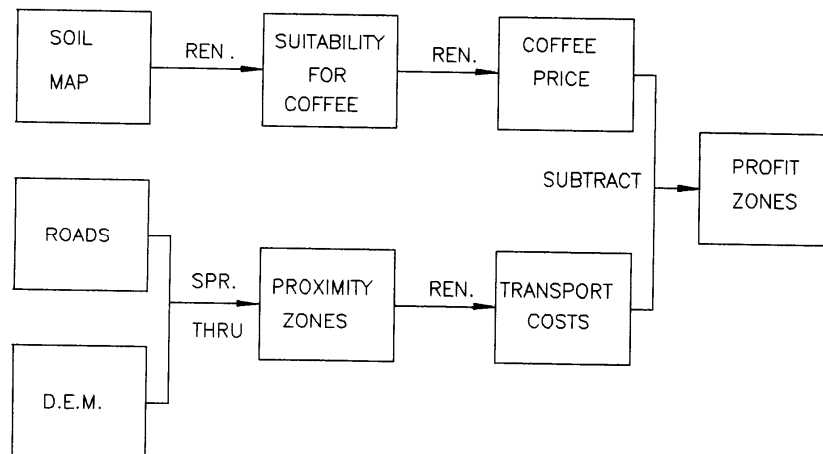
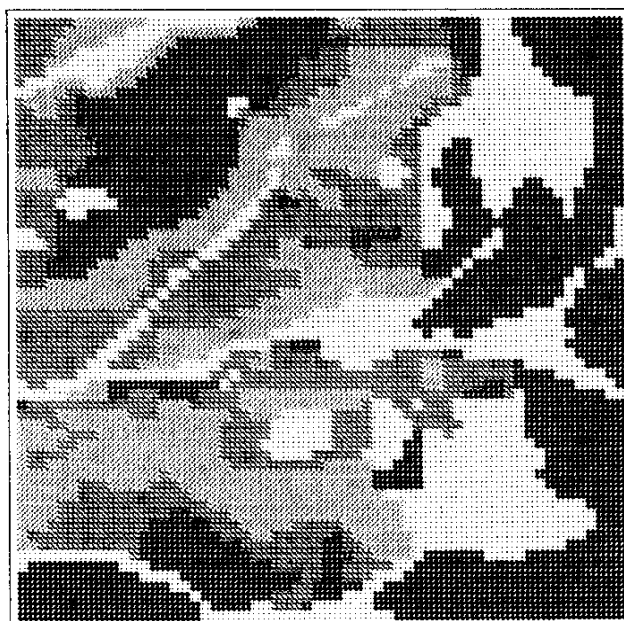


圖 5-12 咖啡收益性評估流程圖



Coffee price

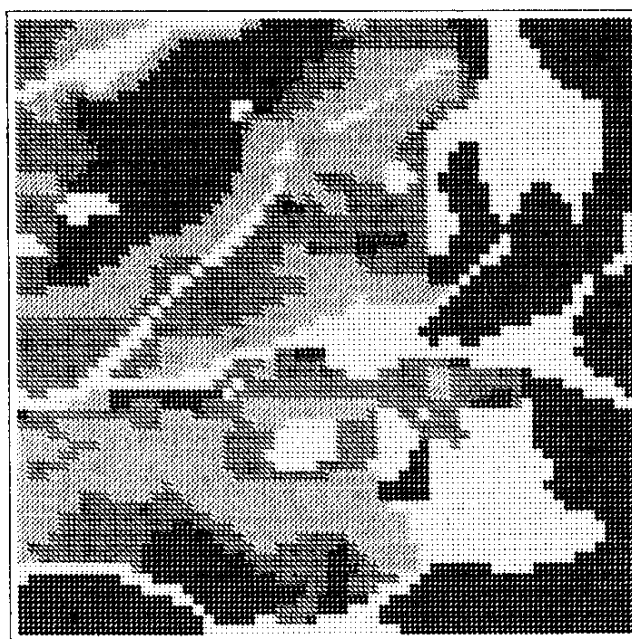
20

40

60

80

圖 5-13 咖啡價



Coffee price

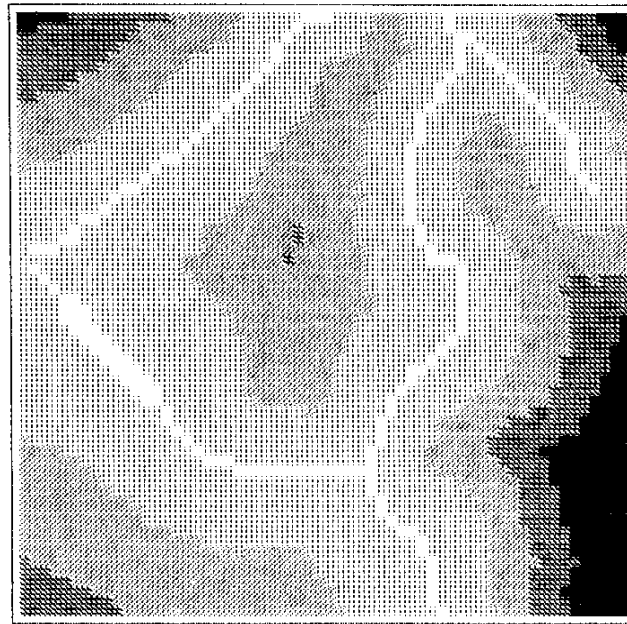
20

40

60

80

圖 5-14 路的最近



Transport costs

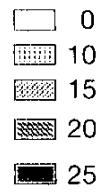


圖 5-15 運輸成本

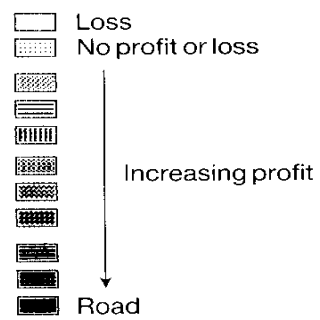
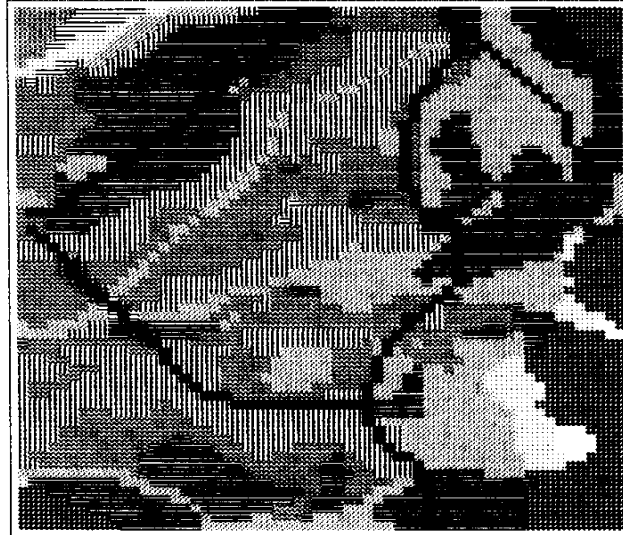


圖 5-16 收益

## **7.地圖模型(cartographic modelling)在土地估算及規劃的優缺點**

優點：使用者可很快的學習如何使用 GIS。

地圖模式(cartographic modelling)的優點：明確地定義問題的需求及決定需要解答的資料。

缺點：cartographic modelling 的主要缺點伴隨著代數及空間算術的決定假設。使用者要了解，對於名詞或排序的變數來說數學運算是沒有意義的。

第二個主要問題是模型步驟隱含假設在疊合時所有編碼的資訊是絕對正確的，沒有任何錯誤的部分。

第三個問題是大部分的空間分析運作是不可交換的，使用者必須特別當心變形發生的順序。

## 第六章 資料品質、誤差及自然變異

GIS 是有用但卻很昂貴的工具，包括硬體、軟體、人員訓練以及資料收集、資料截取及處理的成本。

第 5 章討論隱含的假設(cartographic processes)：模式中所有的資料都無誤差。很少有研究討論 residual variation 及機差如何產生？或在地理資訊化過程中被產生成增大，及這些誤差對於研究結果的效應。本章先對此課題做一回顧，接著詳細闡明誤差增加的一些觀點，表 6-1 說明伴隨著地理資訊過程中，所產生誤差的三組因子，在此，error 表示錯誤及統計變化。

Group I error—明顯易查

Group II error—用資料工作時才查出

Group III error—最難點出，除了資料外還需其他技巧

### I：誤差的明顯來源：

#### 1. 資料的年代：

除了地質學之外，資料年代越久遠，其可信度越低。

#### 2. 地區的疊合：

所需的資料很難一次完全套疊在地圖上，除非比例尺太小，但那又無法滿足我們的需要。如果不能完全疊合，必須決定要達到同一性有多需要？要收集更多資料或產生詳細的資料以吻合較不詳細的資料。

#### 3. 地點範圍尺度：

多數地理資源資料都已產生並以主題圖方式儲存，直到近幾年，數位化資訊系統建立才使得原始野外觀測資料，利用這個過程製成大尺度地圖，不只有較詳細的位相關係(空間解析度)也有較詳細的圖例。地圖的比例視研究課題而定，小比例圖不夠詳細，大比例圖可能有太多資訊使得資料太冗長。

#### 4. 觀測密度：

雖然觀測密度對資料的可信度是一項合理的指標，但並不是絕對的，利用巢巢法取樣技術，以變方分析來估計距離對變方發生增加情形，圖 6-1：地表層(A)：採樣距離自 200m→20m 時，對於累計變方改善不大，其下的圓石黏土層(B<sub>1</sub>, B<sub>2</sub>)則在採樣距離<20m 時才能解決。

簡而言之，樣本密度對資料品質來說是粗略的指標，了解採樣是否在適當的密度對於解決空間型態的課題也很重要。

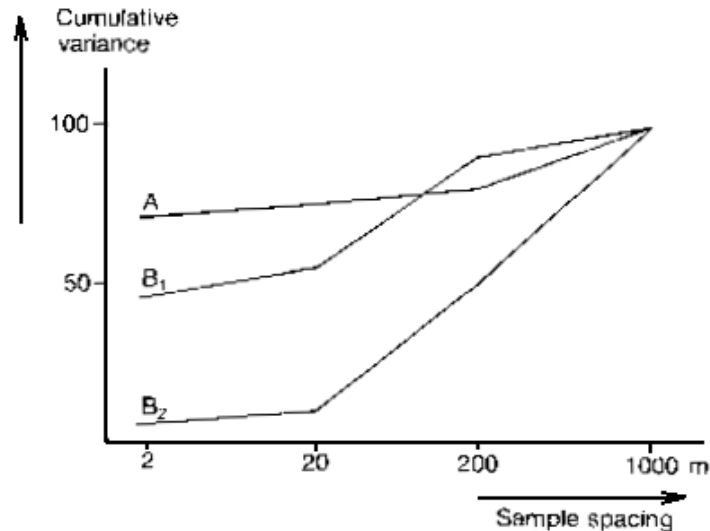


圖 6-1 在某排水區中以築巢法分析土壤變異的結果，其清晰地區分出後更新世土壤的空間結構(A)及其下層的圓石頭黏土。

#### 5.切題(Relevance)：

在地理資訊化過程中，並不是全部資料都可直接切合其應用目的，由於需要的資料並不存在或收集太昂貴而必須選擇代替品(surrogate)如遙感探測的電子信號(electronic signal)。

校正是遙測技術的主要部分，在影像上，一群適當均質的像元被選為訓練樣本，每個頻率寬帶紀錄的反射變化以直方圖的形式展示，其應用是選擇一個像元的訓練樣本(training set)，這個像元的訓練樣本轉回、單一模式分布。這些 training set pixels 可由地真(ground truth)觀測來校正，所以這組的 pixels 等於作物型態，土壤單位或其他可定義的現象。影像內的其他 pixels 被設計成如同 training set 般相同的 set。只因為成本及原始資料測定困難，使用 surrogate 可能是有效的。

#### 6.格式(Format)：

重要的資料格式有三種：

- (1)技術觀點：資料如何寫在磁片上，由一部電腦系統轉移到另一部上，包括電磁介質(magnetic medium)的種類、寫入資料的密度、性質(character)的形態及記錄的長度。
- (2)資料排列的方式：即資料本身的結構。
- (3)資料本身的比例、投影、分類。

總結：資料交換需要將資料重新格式化成較低共通命名的格式，以便許多系統都能讀取，這些格式並不是簡化的，也不是最有效的，但卻是方便的。

#### 7.易取得性(Accessibility)：

並非所有資料都是相同易取得的，土地資源的資料在某國可能可以自由使用，但在另一國家，同樣的資料可能就是機密，成本及格式的問題也會影響資料獲取的難易程度。

## 8.成本：

新資料的收集和輸入或資料的轉變及全新格式都很花錢，在輸入詳細的手繪地圖或把屬性關連到空間資料時數化成本特別高。在等高線及照片影像資料輸入上掃描器可以節省花費。把數化的工作交給專家來做，而把內勤職員調去做更需技巧的工作可能會更便宜。

## II：由於自然變異或原始測定造成的誤差：

### 1.位置正確性：

地理資料的位置正確性之重要性大部分依賴考慮資料的形式，在現代電子量測的技術下，地面上物體的位置可記錄到公分的正確性。反之，土壤或植生單位的邊界常反應了調查者的判斷，如將植生分類成不同類型。

位置誤差由於田間工作做的不夠，原始基本圖的扭曲變形或收縮或網格掃描後向量化品質不佳造成，地區誤差可由畫圖工作站數化的交互作用而校正。一般的位置誤差可由不同種類的轉換而校正，如橡皮伸張法(rubber sheeting)技術。橡皮伸張法的成功，主要依賴資料轉換後的形式，以及轉換的複雜性。當原圖包含了大量的連結的直線片段時，這個方法不是必須的工作。

### 2.內容正確性：

地理資料庫的點、線、面的屬性是正確的或沒有偏差，可分定性正確(如把小麥田當成馬鈴薯田做編碼)及定量正確(如以壞的 pH 值測定器來量 pH 值做推估)。

### 3.資料變異的來源：

(1)量測誤差：來自於不可信賴、不正確或偏差的觀測值或設備。

Accuracy—準確性 precision—精確性

(2)田間資料：在資料品質上，調查者是個關鍵因子。良好設計的資料收集過程及標準可幫助降低觀測者所產生的誤差。

在大量調查組織內，決定及記錄每個調查者的品質是可能的，額外的屬性可以與資料本身存在一起。改善觀測者品質的聰明做法是改善資料收集過程。

(3)實驗誤差：雖然同一實驗室的再現性很高，但不同實驗室都有不同的實驗結果，例如全世界實驗室交換計畫的結果：同一土樣，誤差是粘粒含量  $> \pm 11\%$ , CEC  $\pm 20\%$  (粘粒部份 25%)，BS  $\pm 11\%$ , pH  $\pm 0.2$  單位。

(4)空間變異及地圖品質：許多主題圖，特別是有關於自然性質的，並未計算區域的空間變異來源或在繪圖現象內小範圍改變造成的雜質。這個問題是近來研究的主要課題，研究集中於變異的等級和分類及內插、量化方法可幫助調查及了解，第 8 章將詳細說明。

系統調查的土壤圖估計不會超過 15% 的雜質，雜質(impurities)定義為當特定在地圖 legend 內，觀察不能契合全部需要。

$$Z(\alpha) = \mu + \alpha_j + \varepsilon$$

點 X 的性質 Z 的值

$\mu$ ：圖上全部  $Z$  的平均

$\alpha_j$ ： $\mu$  與單位  $j$  平均之間的差

$\varepsilon$ ：常態分布隨機剩餘變異

實際上，組內變異不足以描述地圖單位內的剩餘變異，因(a)不同地圖單位的相同性質常有不同等級的剩餘變異；(b)剩餘變方的程度經常與地圖面積的大小有權重的關係；(c)地圖單位內性質的變化值，在真實統計上經常不是隨機的，在空間關聯的方法(利用用插法或較大比例調查)可以經常改變。

分類的問題及空間變異將在 7,8 章討論，了解地圖單元內空間變異可構成地圖套疊分析所產生的相對及絕對誤差是很重要的，本章稍後詳細說明如何估算套疊分析所產生的誤差。

### III：過程產生的誤差：

#### 1. 電腦數值誤差：

為記錄數字，電腦字的雙倍準確度(double-precision)對於數學運算及資料儲存，有重要的結果，當程式語言允許變數及陣列被定義，在大電腦完成統計計算時，rounding errors 不像是個問題，比較困擾的是 8 及 16 位元微電腦如表 6-2 所示：1.0000001 平方 27 次的結果。圖 6-2 網格過程資料以整數編碼造成誤差。

$(A+B)+C=A+(B+C)$ 之規則在  $A=1.E10$ ,  $B=-1.E10$ ,  $C=1$  時，當電腦存取少於 10 明顯數值時，規則被達最後一位數需要避免超出準確度。

表 6-2 在某些微電腦中循環產生的誤差：1.0000001 的多次平方結果。

No. of Squares	BBC Micro	IBM personal computer			
	BASIC Single precision	FORTRAN-77 Single precision	BASIC Single precision	BASIC Double precision	BASIC single-double per cent difference
1	1.0000002	1.000000	1.000000	1.00000020000001	0.0000038
2	1.0000004	1.000000	1.000001	1.00000040000006	0.0000077
3	1.0000008	1.000000	1.000001	1.00000080000028	0.0000154
4	1.0000016	1.000000	1.000002	1.00000160000120	0.0000307
5	1.0000032	1.000000	1.000004	1.00000320000496	0.0000615
6	1.0000064	1.000010	1.000008	1.00000640002016	0.0001229
7	1.0000128	1.000020	1.000015	1.00001280008128	0.0002459
8	1.0000256	1.000030	1.000031	1.00002560032640	0.0004917
9	1.0000512	1.000060	1.000061	1.00005120130818	0.0009833
10	1.0001025	1.000120	1.000122	1.00010240523794	0.0019663
11	1.0002050	1.000240	1.000244	1.00020482096271	0.0039312
12	1.0004101	1.000490	1.000488	1.00040968387704	0.0078565
13	1.0008204	1.000980	1.000977	1.00081953559496	0.0157137
14	1.0016416	1.001950	1.001955	1.00163974282851	0.0314297
15	1.0032859	1.003910	1.003913	1.00328217441356	0.0628688
16	1.0065827	1.007840	1.007841	1.00657512149601	0.1257719
17	1.0132088	1.015740	1.015744	1.01319347521471	0.2517049
18	1.0265921	1.031740	1.031735	1.02656101821766	0.5040405
19	1.0538914	1.064480	1.064478	1.05382752412407	1.0106168
20	1.1106872	1.133110	1.133113	1.11055245060146	2.0314411
21	1.2336261	1.283940	1.283944	1.23332674553692	4.1041535
22	1.5218334	1.648510	1.648513	1.52109486125669	8.3767437
23	2.3159770	2.717600	2.717595	2.31372957694150	17.4551843
24	5.3637495	7.385320	7.385325	5.35334455521391	37.9572042
25	28.7698096	54.543000	54.543020	28.65829792683842	90.3219043
26	827.701947	2974.940000	2974.941000	821.29804006343120	262.2242807
27	685090.513	8850270.000000	8850273.000000	674530.47061203350000	1212.0642262



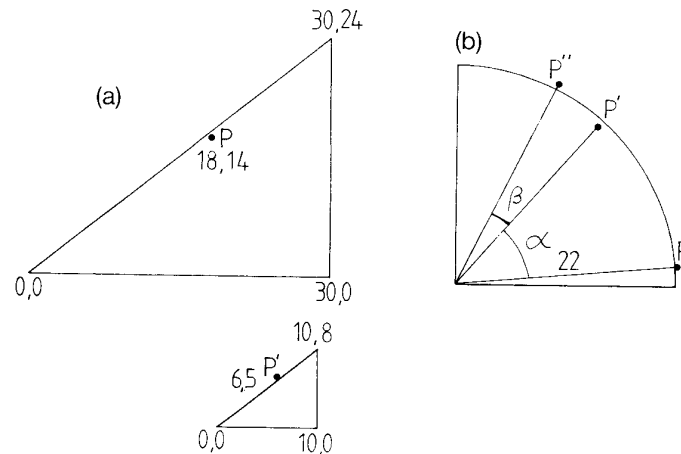


圖 6-2 當存在數學整數和網格是太粗糙時，循環測量(a)和旋轉(b)會導致位相誤差

表 6-3 電腦字元長度和數化範圍與精確度的關係

Number of bits	Number of significant digits (decimal)	Approximate decimal range
16 integer	4	$-32768 \leq x \leq +32767$
32 integer	9	$-2 \times 10^9 \leq x \leq 2 \times 10^9$
64 integer	18	$-9 \times 10^{18} \leq x \leq 9 \times 10^{18}$
80-bit packed decimal	18	$-99 \dots 99 \leq x \leq +99 \dots 99$ (18 digits)
Short real 32 (single precision)	6-7	$8.43 \times 10^{-37} \leq x \leq 3.37 \times 10^{38}$
Long real 64 (double precision)	15-16	$4.19 \times 10^{-307} \leq x \leq 1.67 \times 10^{308}$

電腦字長度造成的問題可能解決之道：

- (1)以雙倍準確字工作
- (2)以 32 位元整數工作

(3)最有用的 nested approach：把地區分成幾個解析度較粗的 tile，在每個 tile 參考原始局部地區，詳細的測定全部座標。Nesting 可視需要反覆進位，因為一個 tile 含有使用者任何時候卻可在螢光幕上看到的許多資料，當使用局部過程時，nesting 不是限制因子。

2.聯合位相分析的缺點：

應用在地理資訊過程化中隱含了下列假設：

- (a)資料來源是單一的；
- (b)數化過程是永不出錯的；
- (c)地圖疊合只是邊界交叉及線網路全新接合的問題；
- (d)邊界可以強烈的定義及畫出；
- (e)數字可決定一切；
- (f)對於地圖屬性來說，用一個或其他自然理由定義等級間隔是最好的。

### 3.分級輸入及一般化：

資料輸入 GIS 不規則性，可由分類及從點資料到面的內插屬性化的方法，電腦套疊和邊界交叉所產生問題和錯誤。在下節，討論以下幾點造成的可能誤差：

- (a)在網格中以一組網格單元描繪向量多邊形網。
- (b)由邏輯和數學在 2 個或疊合 grid cell 或多邊形網路上運算。
- (c)由疊合及交叉 2 個多邊形網路。

最後，考慮利用邊界描述自然現象之空間變異的問題。

#### 1.向量式地圖網格化造成的誤差：

當位相圖由網格單元來表示時，誤差會以兩種方式產生：

- (1)最常見的是每個網格包含一個屬性值，是 cell 的平均值。
- (2)當多邊形地圖接近一個 grid 時，分類上的誤差就會產生。

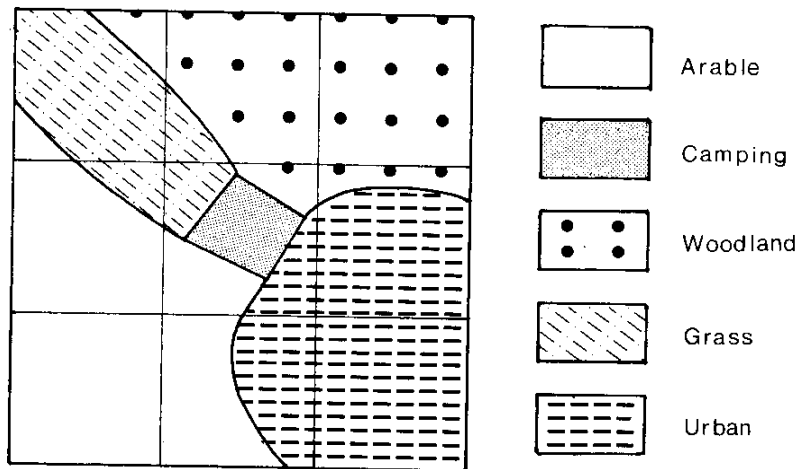


圖 6-3 網格單元太大以致於無法解析空間詳細編碼問題，  
如何將中央之網格單元予以編碼？

#### Switzer's method :

目的：估計網格圖的準確度。

假設：誤差只是利用 grid cell 中央一系列的點估計原圖的結果。

本法適於理想的地誌圖，亦即小區間區別 明顯。

過程：圖 M 分成 K 個均一的單位或顏色，每部分稱為一個小區，分為 n 個樣區，其排列規則一致，由 cell 中間點來定義該 cell 性質字為  $M_1, M_2, \dots, M_k$ ，設總面積等於單位， $A(M)=1$ 。

地圖估計或網格過程的誤差：

$$L_{ij} = A(M_i \cap M_j) \quad i < j \quad L_{ij} : \text{真正是 } i, \text{但在估計上是 } j \text{ 的面積}$$

$$L_i = \sum_{j > i} L_{ij} \quad (\text{固定 } i) \quad M_i : \text{不正確部分的總面積}$$

$$L = \sum_{i=1}^k L_i \quad \text{全圖中接合不良總面積}$$

接合不良的程度是 2 個互相獨立因子的函數：(a)真實地圖的複雜性(b)採樣網的幾何性質。

$P_{ij}(d)$ —真實地圖單位  $i$  內隨機的點及真實地圖單位  $j$  內 cell 中央的點間隔  $d$  時的機率，可當做無限的，可微分的函數，可用泰勒展開式來表現。

$$P_{ij}(d) = P_{ij}(0) + P'_{ij}(0)d + P''_{ij}(0)d^2/2 + \dots$$

$$\because P_{ij}(d=0) = 0 \text{ 且 } P_{ij} \geq 0, d \geq 0 \therefore P'_{ij}(0) > 0$$

$$L_{ij} = P'_{ij} \Sigma A(Sh) Dh$$

$A(Sh)$ : cell  $Sh$  的面積， $Dh$ : 採樣 cell  $Sh$  內隨機一點與 cell 的資料點當全部 cell 有相同形狀，且二者間平均距離，資料點都在 cell 中間時，

$$L_{ij} = P'_{ij} \times D$$

接合不良的估計包括了  $P'_{ij}$  及  $D$  的估算。

$D$  的估算：矩形網

$$D = n^{1/2} \times \left( 2\sqrt{r + \frac{1}{r}} + \frac{1}{\sqrt{r^3}} \ln(r + \sqrt{1-r^2}) + 2\sqrt{r^3} \ln\left(\frac{1}{r} + \sqrt{1 + \frac{1}{n}}\right) \right) / 12$$

$r$ : cell 邊長比值

$$r=1 \text{ (正方形)} \quad D=0.383n^{-1/2}$$

$$r=0.6 \text{ (標準印表機 長方形)} \quad D=0.417n^{-1/2}$$

⋮

結論：正方形格子，資料點位在 cell 中央時，與最有效的六邊形採樣比較，只多了 1.6% 誤差，當使用 cell 內隨機位置的資料點時，增長 cell 造成較大的接合不良。

$L_{ij} = P'_{ij} \times D$  對比較採樣設計的效率適合的，但對真正的接合不良估計要再加  $d^2$  項而可改良。

$$L_{ij} = P'_{ij} \Sigma A(S_n) D_n + \frac{1}{2} P''_{ij} \Sigma A(S_n) D_n *$$

$$\text{簡化: } L_{ij} = P'_{ij} D + \frac{1}{2} P''_{ij} D *$$

正方形格採樣點在 cell 中央時， $L_{ij} = 0.383n^{-1/2} P_{ij} + 0.083n^{-1} P''_{ij}$

$$P'_{ij} = \frac{1}{2} n^{\frac{1}{2}} \left\{ 4P_{ij} n^{\frac{-1}{2}} - P_{ij} \left( 2n^{\frac{-1}{2}} \right) \right\}$$

$$P''_{ij} = \frac{2}{3} n^{\frac{1}{2}} \left\{ 2P_{ij} n^{\frac{-1}{2}} - P_{ij} \left( 2n^{\frac{-1}{2}} \right) \right\}$$

$$L_{ij}(\text{二次}) = 0.60 P_{ij} \left( n^{\frac{-1}{2}} \right) - 0.11 P_{ij} \left( 2n^{\frac{-1}{2}} \right)$$

$$P_{ij} n^{\frac{-1}{2}} \text{ 及 } P_{ij} \left( 2n^{\frac{-1}{2}} \right) \text{ 機率的估算:}$$

1. 在距離  $d=1$  cell 寬估計 cell 對的總數  $NPAIRS = 4PQ - 2d(P+Q)$

$P$ : 行數  $Q$ : 格內列數 第二項是校正項

2. 每一繪圖單位  $ij$  對，沿著格子上下不同地圖單位格子算 cell 對數(TALLY)

3. 計算  $P_{ij}(n^{-1/2}) = \text{TALLY}_{ij} / \text{NPAIRS}$

4.  $d=2$  時，重覆 1-3 步計算  $P_{ij}(2n^{-1/2})$

5. 由  $L_{ij} = 0.60P_{ij}(n^{-1/2}) - 0.11P_{ij}(2n^{-1/2})$  計算  $L_{ij}$

6. 把每一繪圖單位  $L_i$  當做  $L_{ij}$  的總和，計算全部接合不良，注意  $L_{ji} = L_{ij}$

7. 由  $L_i$  的總和計算全部接合不良。

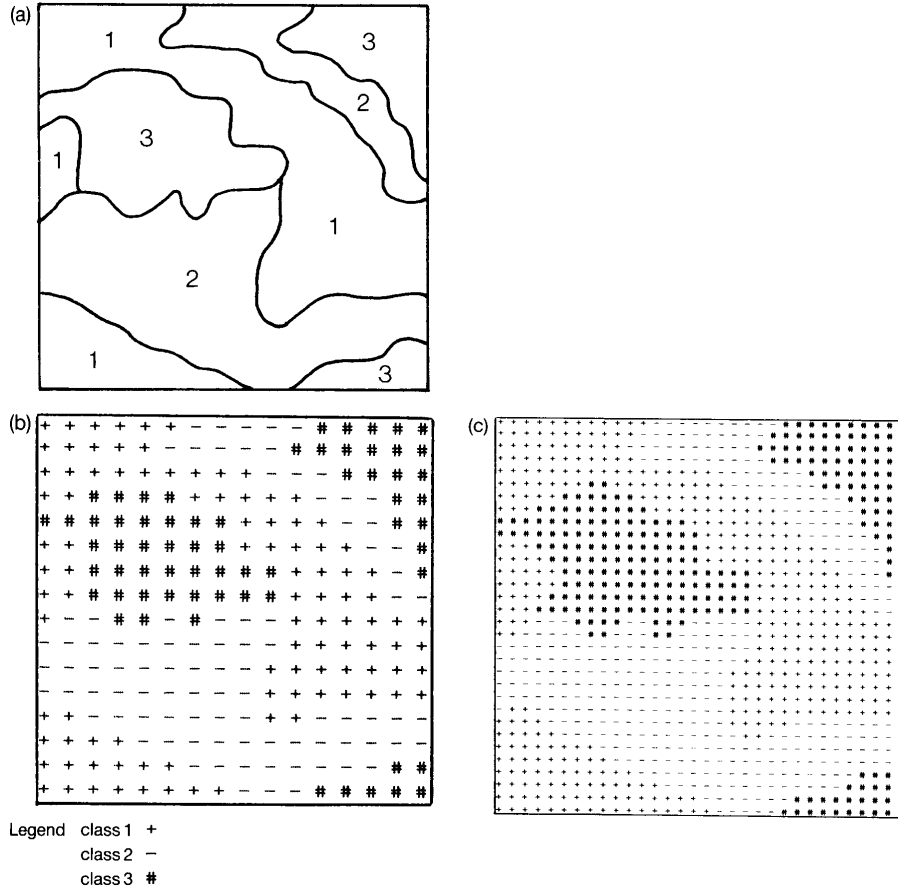


圖 6-4 (a)三種區域的原始向量多邊形圖(b)同圖以 16×16 方格數化(c)同上，但以 32×32 方格數化

結果：表 6-4(a)vector 資料以 6-4(b)、6-4(c)網格表示

	(16×16)	(32×32)
	↓	↓
total mismatch %	→8.46	4.11
Mismatch estimates using quadratic fit on square grid		
(a) 16 × 16 grid		(b) 32 × 32 grid
Polygon pairs	Probability estimates	
	$P_{ij}(n^{-1}/d)$	$P_{ij}(2n^{-1}/d)$
1 2	0.05104	0.02545
2 3	0.02188	0.01184
3 2	0.03438	0.01689
	Mismatch per polygon pair	
	L12 0.020 L21 0.020 L31 0.008	L12 0.010 L21 0.010 L31 0.004
	L13 0.008 L23 0.014 L23 0.014	L13 0.004 L23 0.006 L23 0.006
	Mismatch per polygon	
	L1 0.028 L2 0.034 L3 0.022	L1 0.014 L2 0.016 L3 0.011
	Total mismatch % = 8.46	Total mismatch % = 4.11

2.地圖數化或地圖編碼(geocoding)時產生的誤差：

潛在的誤差有 2 個來源：(a)原因(b)數化表現

(a)畫出的界線太粗，如 1mm 寬的線在 1：1,250 圖上就是代表 1.25m，在 1：100,000 就是 100m。

(b)曲線形狀與頂點數目有關，同時，數化直線的相對誤差遠低於數化複雜的曲線。

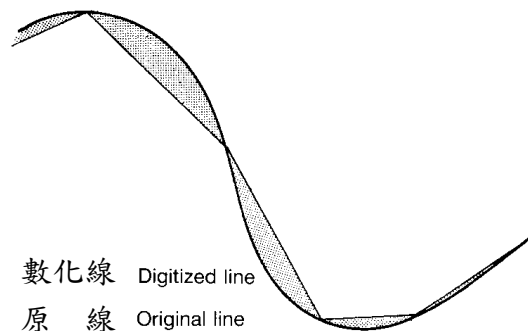


圖 6-5 數化一條曲線是一取樣過程

不準確的邊界總長(Total boundary inaccuracy)：

$$L_i = \sum_{i=1}^N (h_i * l_i) / T$$

當界線是同形時， $H = (h * L) / T$

線總長度  $L = (T, K) / 0.6366$

$h_i$ ：線  $i$  的水平誤差

$l_i$ ：線  $i$  的長度

$N$ ：地圖上界線數目

$T$ : 地圖上總面積

$K$ : 交點數目

Epsilon distance(製圖線以其目的物的一個概括平均距離): 點  $x, y$  位於多邊形  $A$  內有下列位置:

(i)絕對之內, 圖 6-6 之點 7

(ii)絕對之外, 圖 6-6 之點 3

(iii)可能在內, 圖 6-6 之點 2 possible in/out 2polys—4

(iv)可能在外, 圖 6-6 之點 1 possible in/out > 2polys—5

(v)不確定(在界線上), 圖 6-6 之點 6

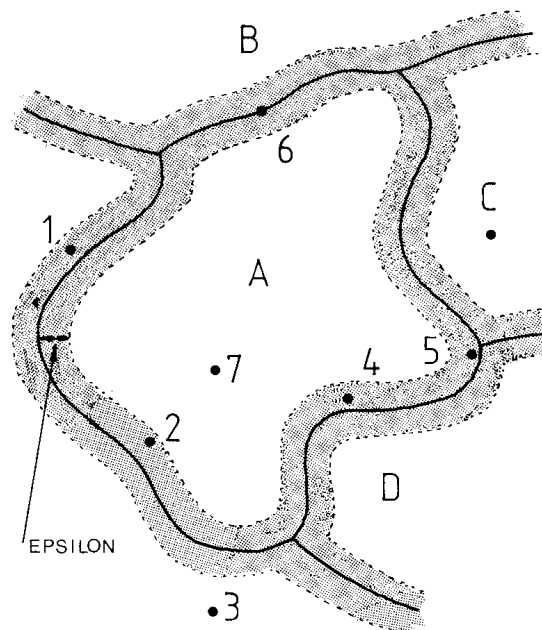


圖 6-6 誤差帶(epsilon 寬)和點內多邊形之不確定位置標示(接錄自 Blakemore 1984)

### 3. 疊合 2 個以上多邊形網造成的誤差:

由於數個多邊形疊合, 將會有  $n$  個線段,

$$n = \left[ \sum_{i=1}^k M_i^{1/2} \right]^2$$

疊合得到的圖中增加了許多小的多邊形, 利用大小及邊界複雜性把得出的多邊形分類來計算疊合的結果。

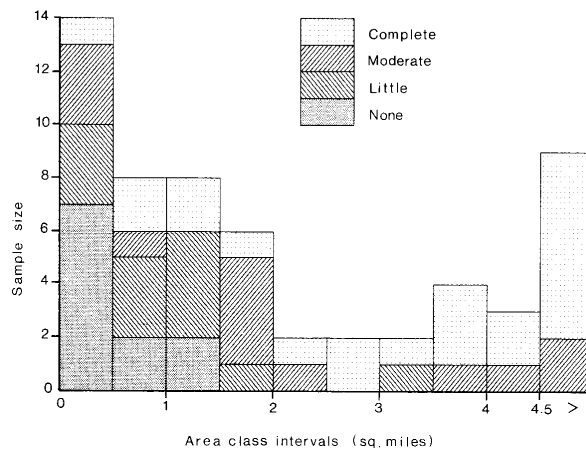


圖 6-7 研究測試三種多邊形網路套疊後的結果

Goodchild(1978)討論多邊形套疊問題：認為所得到多邊形(套疊後)的數目主要是受到邊界複雜性影響，其次才是多邊形數目。

多邊形邊界套疊後，嚴重的問題是產生許多小的假多邊形(spurious polygon)，假多邊形包含了2個互相矛盾的現象，(1)個別地圖上邊界數化越正確，座標越多的，假多邊形產生的數目越多，(2)畫地圖的主要方法，在疊合時用人工方法設計來幫助眼睛，在數化地圖工作上造成較大的問題。

說明：spurious polygon 最常出現在相同線段數化二頂點的連接上。

$n_1 * n_2$ ：端點數  $S_{min}=0$

S：spurious polygon 數目  $S_{max}=2\min(n_1, n_2)-4$

期望值  $E(s)=(2n_1*n_2)/(n_1+n_2)-3$

實際上，spurious polygon 數目從來超過  $S_{max}$  的 71%，spurious polygons = 多邊形網格化時，較大的問題是把他們移除後避免影響到最後的圖。

解決：考慮距複雜線段某定距離所有的點，做為斷線位置的估計，然後用最小平方法或最大概似法來配合新線，對於土壤圖(邊界不明確)及行政區圖(邊界明確)的疊合可以用加權位相邊界線(topographic boundary)方式進行。

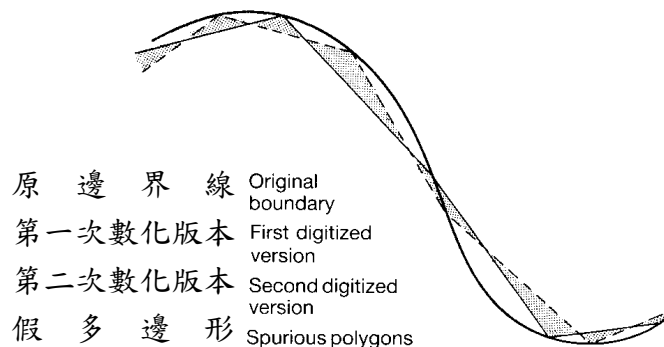


圖 6-8 假多邊形如何在地圖套疊內發生，注意，在這例子中，以適配正常多邊形或 Spline 功能的數化點，將假多邊形能完全被消除

#### 4.自然的邊界線

邊界性質：

行政界線—要明確

土系界線—不宜明確

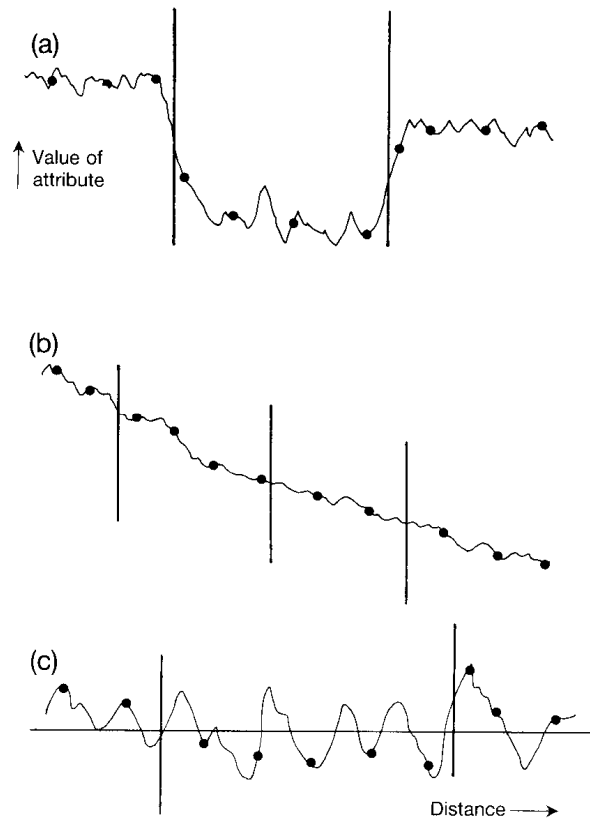


圖 6-9 當製作天然資源圖時遭遇的三種邊界線

(a)不連貫：很好切

(b)連續趨勢：不好切

(c)跳動，邊界無意義，皆為小範圍變化  
• 為樣本位置

5.邊界在統計上的性質：

Webster and Wong (1969)發明一個方法來找一維樣品的邊界

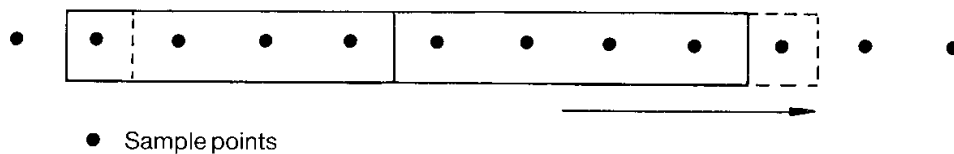


圖 6-10 比較二組點之樣品族群是否差異

$$t = \frac{|X' - X''|}{S} \times \left( \frac{n' n''}{n' + n''} \right)^{\frac{1}{2}}$$

X'X''：兩半的平均值

n'n''：兩半的點數

S：標準偏差



t test 分析隨 window 寬度而異，越窄，邊界越多。

當逐漸改變時，這種 simple split window 無法找到適當邊界位置。

Webster(1978)利用 window 上 mullion 刪除近中間的樣品以增加兩半的對比。

以上皆為一維，二維可用 local windows or kernels。

碎形 (Fractal)：

Degree of boundary 與 overlay、clean-up 過程併入一起。有 2 個問題：(1)2 個土壤單位分界的顯著程度長度而變，在進行時會過度完整。(2)實際位置及邊界長度端賴地圖比例尺度；海岸線的估計長度會隨地圖比例尺的增大而增大，這是因為地圖比例增大時，許多原先無法解決的地理特徵都可畫出輪廓。

可解決的詳細程度是比例尺的函數，增加地圖比例尺並不會絕對增進精密度，只是顯示迄今仍未注意到的變異，碎形有 2 個重要特徵：(1)"自我類似"，在一個比例尺變異會重複出現在另一個，(2)部分維度的特殊性質。

一條碎形曲線(A fractal curve)：是一個連續線系，如多項式，當比例尺增加時，越看越像直線，在這條線上任選 A, B 二點，任意二點間距離端賴測定之比例尺，當比例尺改變時，其距離並不會改變，在較大比例尺下檢查 fractal sense 時，造成解析度更加詳細。

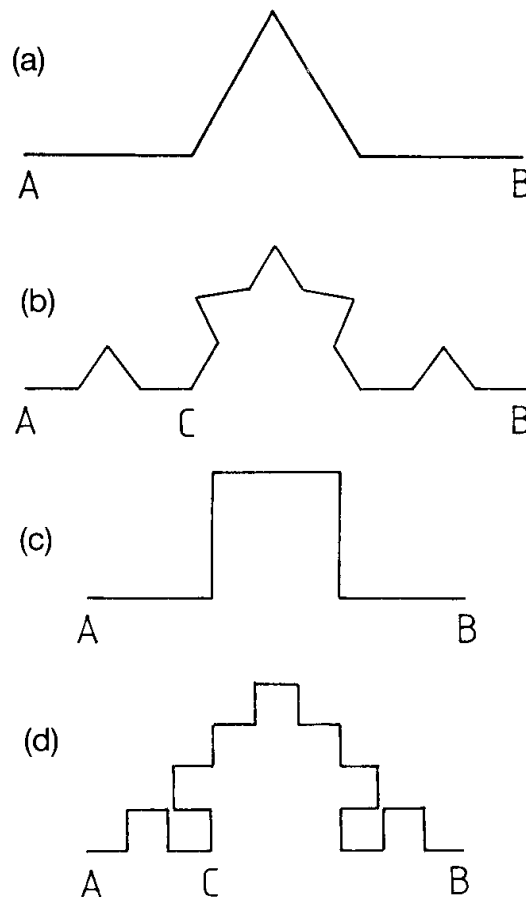


圖 6-11 碎形曲線詳細情形：(a)、(b)  $D=\ln 4 / \ln 3 \doteq 1.2618$ ；(c)(d)  $D=\ln 5 / \ln 3 \doteq 1.4650$

Fractal dimension  $D=\log N / \log r$

N：測定某型單位長的步驟數

r：比例尺比率

D 值：liner fractal curve：1-2

D=1：曲線平滑，可完全接近多項式，在任一點可找出其切線

D>1：線有一 fuzziness band

D=2：fuzziness 很大，成為一個區域

表面上，d 範圍 2(完全平滑)~3(無限皺)

self-similarity：如地理上的物體(山、河)是真的 fractals，交側的變異可以隨比例尺化，反之，土地形態變化在幾米所見和幾百米或幾千米在相同尺度參數下統計上皆相似，也有些地理學家反對：

Poisson dist

Gapninar dist

Weibull dist

估計碎形維數(Estimating the fractal dimension, D)：

D 的估算，不同方法 summary T-6-6→T-6-7

土地形態變化流體(如水、空氣)有較小 D 值，亦即其改變是平滑，長範圍變化為主。反之，土壤性質、礦物岩石的變化則是小範圍的改變，這些結果與我們的認知很符合，亦即土地形態、地下水面、空氣壓力應用連續地圖(isopleth)來模型化，the fuzzy 土壤性質不明確的變化，花太多注意力在土壤邊界的確實位置或繪圖的高度確性在邊界上，尤其是人工的正確造成地圖接合與疊圖技術上的問題更是不智。

用一組資料來估算 D 值，會隨著不同比例尺度而改變測值，通常在相同景觀下，D 值會隨高度而增加，可能因為在低海拔沉積過程而有較平坦的土地形態，而在高山上則有風化及沖蝕過程。

邊界不是絕對的，它倒是加工品或中間折衷產物，它們的統計(長、寬)是畫地圖及比例尺性質變異的複雜函數。

由疊圖結合屬性：

(a)數學上的關係： $u = f(X_1, X_2, X_3, \dots, X_j, S)$

$$Su = \sum_{i=1}^I \sum_{j=1}^J \left\{ \frac{\partial u}{\partial_i i} \times \frac{\partial u}{\partial X_j} \times SX_i \times SX_j \times \dots r_{ij} \right\}$$

和或差：無相關：

$$u = X \pm Y \pm \dots$$

$$Su = \sqrt{SX^2 + SY^2 + \dots}$$

$$\text{例：} X=10 \pm 1 \quad Y=8 \pm 1 \quad \therefore u=10+8=18 \quad Su = \sqrt{1+1} = 1.4$$

相對機差  $1.4/18=8\%$

變數有關聯的加法：

$$u = X + Y \text{ 時, } Su = \sqrt{SX^2 + SY^2 + 2SX * SY * r_{XY}}$$

$$\text{例：} u=(10 \pm 1)+(8 \pm 1) \quad Su = \sqrt{1+1+2*1*r_{XY}}$$

積或商—無相關

$$u = X^a \times Y^b \quad \delta u / \delta X = aX^{a-1} \times Y^b \quad \delta u / \delta Y = bX^a \times Y^{b-1}$$

$$Su = \sqrt{a^2 X^{2(a-1)} Y^{2b} SX^2 + b^2 X^{2a} Y^{2(b-1)} SY^2}$$

$$\text{例：} u = 8 \times 10 = 80 \quad Su = \sqrt{8^2 \times 1 + 10^2 \times 1} = \sqrt{164} = 12.8$$

相乘積，除了絕對機差外，也相對機差  $12.8/80=16\%$

$$u = aX \text{ 時} \quad Su = \sqrt{a^2 SX^2}$$

$$u = AX^a \text{ 時} \quad Su = \sqrt{A^2 a^2 X^{2(a-1)} SX^2} \quad r = 1$$

$$\text{例：} X = 10 \pm 1 \quad u = X^2 = 100 \quad Su = \sqrt{(2X)^2 SX^2} = \sqrt{20^2 \times 1} = 20$$

相對機差  $20/100 = 20\%$

對數及其他關係：

$$u = B \ln X \quad \delta u / \delta X = B / X \quad Su = \sqrt{B^2 / X^2 \times SX^2} = \frac{B}{X} SX$$

error 增減端賴  $B/X$

$$u = A \times \sin X \quad Su = A \times SX \times \cos X$$

數學運算推薦策略：

- (a) 避免變數之間有相關
- (b) 可能的就加
- (c) 不能加，就用乘除
- (d) 盡量避免相減或增加變數乘方

我們可以利用更正確的測定方法或更加的內插技術來降低地圖製作變數的變方。

實例說明：

$$A = R \times K \times L \times S \times C \times P$$

$$A: \text{土壤年溼失量(T/ha)} \quad \text{算出 } A = 23 \pm 14.8 \text{ t/y}^2$$

$$R: \text{降雨造成沖蝕量} \quad R = 0.11abc + 66 = 297 \pm 72$$

a: 平均年降雨量

b: 2 年發一次最大年降雨量(cm)

c: 在 2 年內發生的一年降雨最大總雨量(cm)

$$K: \text{土壤沖蝕力} \quad K = 0.1 \pm 0.05$$

$$L: \text{坡長(m)} \quad L = (\ell / 22.1)^{1/2} = 2.13 \pm 0.045 \quad \ell: \text{坡長(m)}$$

$$S: \text{坡度(\%)} \quad S = 0.0065s^2 + 0.0454s + 0.065 = 1.169 \pm 0.122 \quad s: \text{坡度(\%)}$$

$$C: \text{耕種參數} \quad C = 0.63 \pm 0.15$$

$$P: \text{保護測定} \quad P = 0.5 \pm 0.1$$

(b) 邏輯關係(AND, OR, NOT)

畫地圖定小區域時，區域中間是 well-defined，而週邊則是 less well-defined。

Membership function: in fuzzy set theory, 用來描述一個 cell 或面積是這組成員的程度，利用 fuzzy set approach, 允許用語言學上較不精確的字眼，如 most,

some, a few 在地圖疊合問題。

地圖 X 及 Y 疊合成 U，在圖 U 上單一 cell, CELL，如果 A 是地圖上某一等級， $f_A(CELL)$  表示分布在 cell CELL A 的權重。

Fuzzy distance between CELL' and CELL"  $\rightarrow$

$$d(CELL', CELL'') = \sum [f_X(CELL') - f_X(CELL'')]$$

$$d(CELL', CELL'') = \sqrt{\sum [f_X(CELL') - f_X(CELL'')]^2}$$

$A$  in overlay  $X$ ,  $B$  in overlay  $Y \Rightarrow$

$$f(A \text{ and } B)(CELL) = \min(f_A(CELL), f_B(CELL))$$

$$f(A \text{ or } B)(CELL) = \max(f_A(CELL), f_B(CELL))$$

$$f(\text{not } A)(CELL) = 1 - f_A(CELL)$$

結論：

本章回顧了資料機差及變方主要來源，在原始資料的自然空間變異可能是看不見的機差之主要來源。當前主題圖之應用，在平滑線下及同一顯現繪圖單位中隱藏了自然的變異。

改善環境資料收集、分析及製圖的策略，包括：

- (a) 較佳及更適當分類技術；
- (b) 改善內插的技術，以避免用平均及再現位址做為大面積描述的基準；
- (c) 降低田間觀測的機差。

有些論點的解答在下面幾章討論。

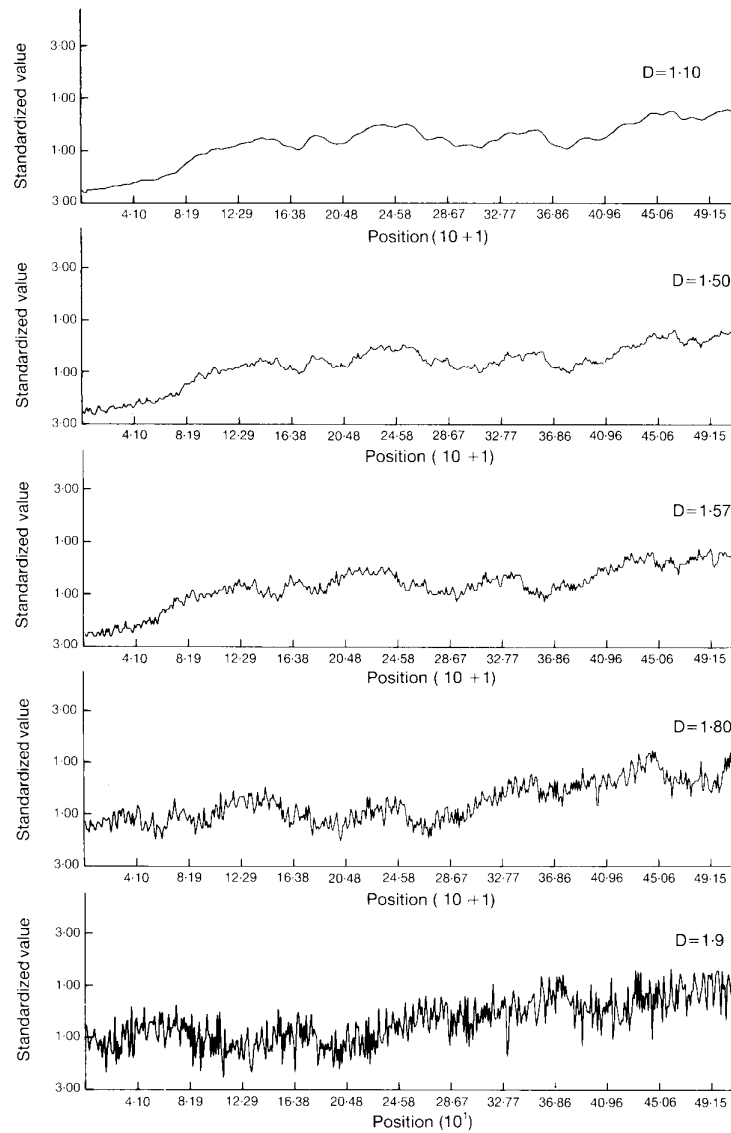


圖 6-12 不同  $D$  值的線性碎形 Brownian 公式，本圖以第一次數化樣版的 512 點模擬 1024 點的系列圖

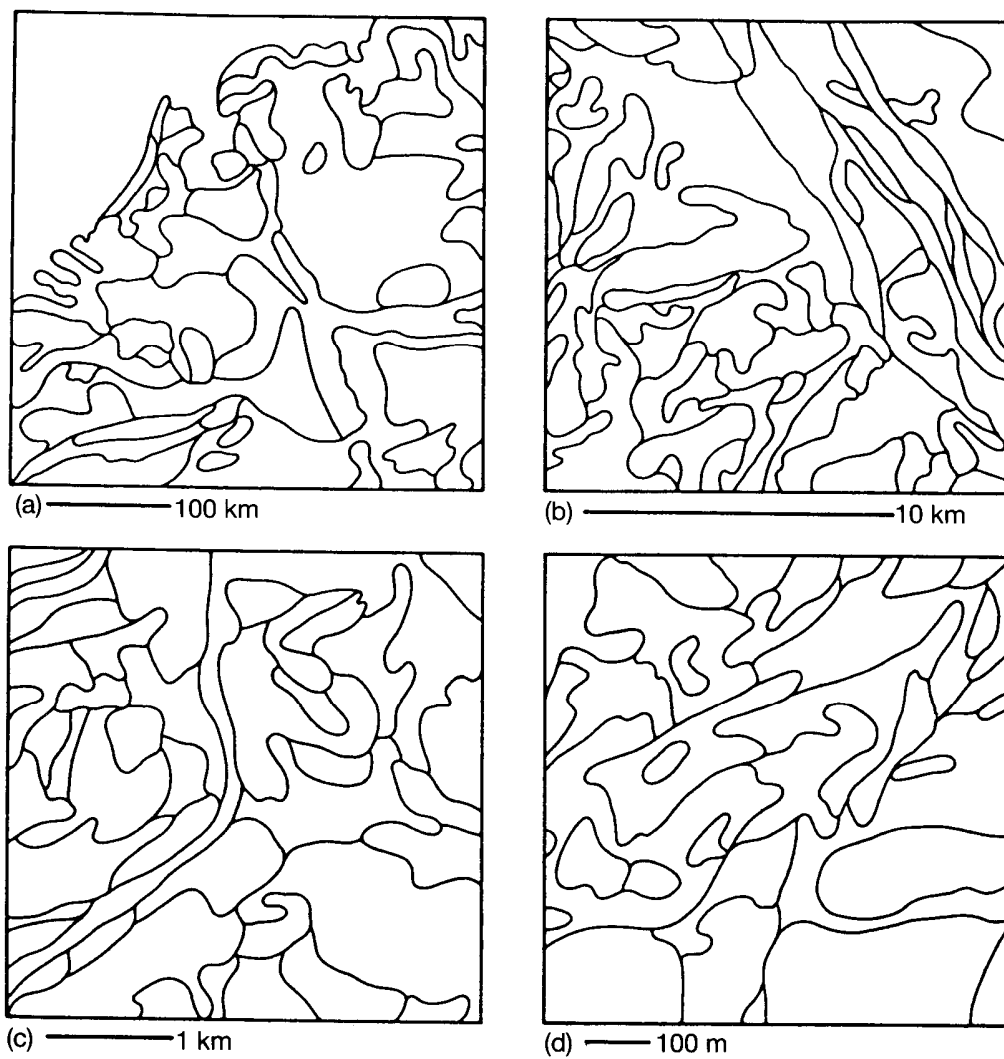


圖 6-13 西北歐不同尺度的土壤分布格局

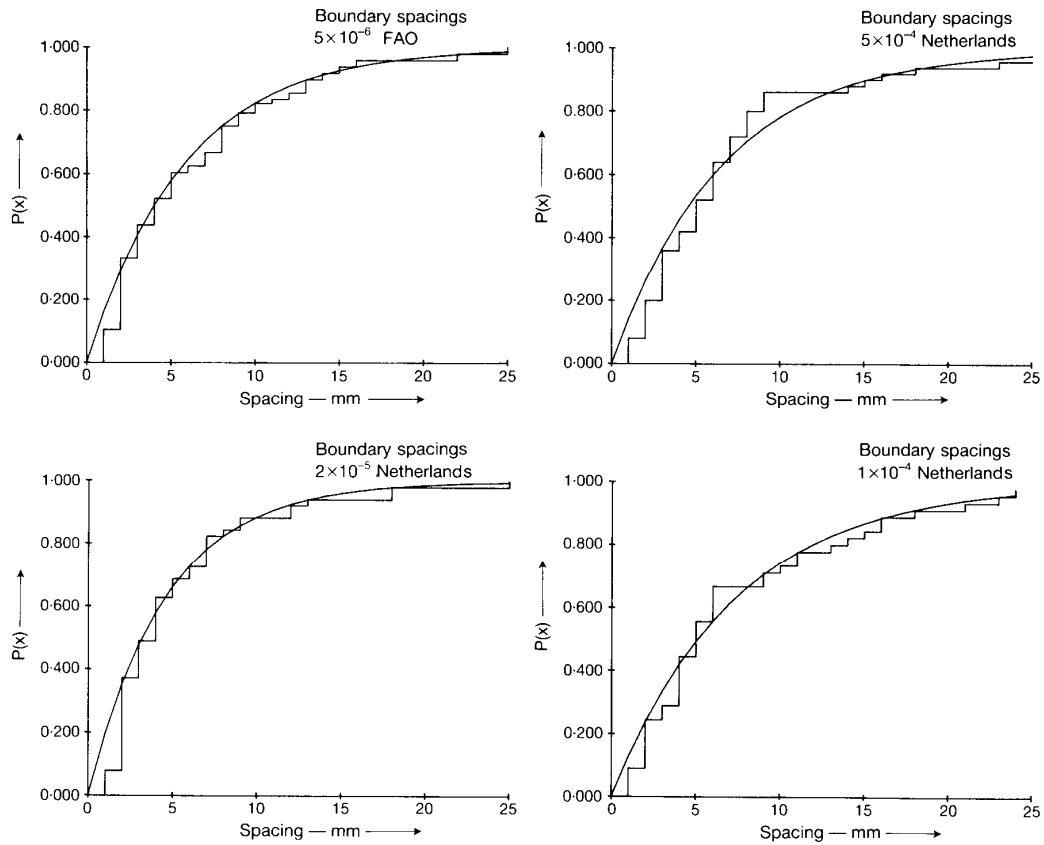


圖 6-14 圖 6-13 的邊界線空間化所得的累積機率曲線圖

表 6-5

Category	Per cent
1. Possibly out	1.5
2. Possibly in	4.4
3. Unassignable	1.4
4. Possibly in/out 2 polys.	29.8
5. Possibly in/out > 2 polys.	6.7
6. Ambiguous	1.2
Subtotal	45.0
7. Definitely in	55.0
Total	100.0

表 6-6 地理資料用於推估 D 值的方法彙整表

Name of method	Relation	Estimate of $D$
1. Length of a trail	$L(\lambda) = k\lambda^{1-D}$ $\lambda$ is step size $k$ is constant	Plot $\log L$ against $\log \lambda$ Slope of line is $1-D$
2. Area/perimeter relationship	$A = kP^{2/D}$ $A$ is estimated area $P$ is estimated perimeter $k$ is a constant	Plot $\log A$ against $\log P$ Slope is $2/D$
3. Variance of increments or semivariance (one-dimensional transect)	$2\gamma(h) = h^{(4-2D)}$ where $h$ is sampling interval	Plot $\log \gamma(h)$ against $\log h$ Slope is $(4-2D)$
4. Power spectrum (one-dimensional transect)	$P(\omega) \approx \omega^{-(5-2D)}$ where $P(\omega)$ is the power and $\omega$ is frequency	Plot $\log P(\omega)$ against $\log \omega$ Slope is $-(5-2D)$
5. Rescaled range $R/S$ analysis	The ratio of the rescaled range $R(t,s)$ for a series over time $t$ , to the sequential variance $S^2(t,s)$ is proportional to the lag $s$ .	Plot $\log R/S$ against $\log s$ Graph has a slope of $(2-D)$
6. Korcak empirical relation for islands.	$Nr(A > a) = F'a^{-(D/2)}$ where $Nr(A > a)$ is the number of islands above size $a$ , $F'$ is a positive constant.	Plot $\log Nr(Aa)$ against $\log a$ ; slope is $-(D/2)$

NB Full definitions of these methods will be found in Mandelbrot (1982).

表 6-7 一些地理資料推估的 D 值

Description	Apparent $D$ (linear)
Varve data	1.02-1.5
Tree ring indices	1.22-1.45
Annual precipitation	1.13-1.36
River flows	1.02-1.52
Thickness and lithology of Palaeozoic sediments	1.25-1.45
Iron ore in rocks	1.4-1.9
Groundwater levels	1.3-1.8
Landform	1.1-1.3-1.7
Soil properties	1.5-2.0



## 第七章 分類方法

資料分析 { structure or signal : 我們可了解的  
              { noise : 我們無法掌握的

下二章目的：描述一些方法，可讓使用者比用統計和數值方法在傳統地圖上由原始資料得到更多資訊。在進行時，應該改善用分析，模型化結果出現的機差。  
分類：

一個單一連續變數的值的範圍分類：

簡單分類常用來展示連續變化的空間性質，如人口密度、土地形態梯度、空氣污染程度及其他易了解之地誌圖，Evans 分類：

exogenous：區間是固定的，依照開始值與研究資料組有關。

雖然他們是標準分類系統之一部分，他們並不需要成為問題中最被指定的面積，exogenous class intervals 廣泛應用於土壤調查及土地估算分類系統。優點是應用性很普遍，最大缺點是不能解決在任一特定面積的變異。

Arbitrary：無特定目的時被選用，常常其區間是不規則分布，資料也未經過適當檢查。

Idiographic：被選用與資料組特別目的有關，其方式是把資料分成許多組。

Serial：限制在直接數學關係，畫圖的變化是在統計上描述：

(a)常態分布：平均 $\mu$ 、標準偏差 $\delta$ ；

(b)以偏差標準的比例來區分，中心點是平均值；

(c)相等的數學的區間，分類寬度相同；

(d)在倒數：三角、幾何、對數上成等間距。

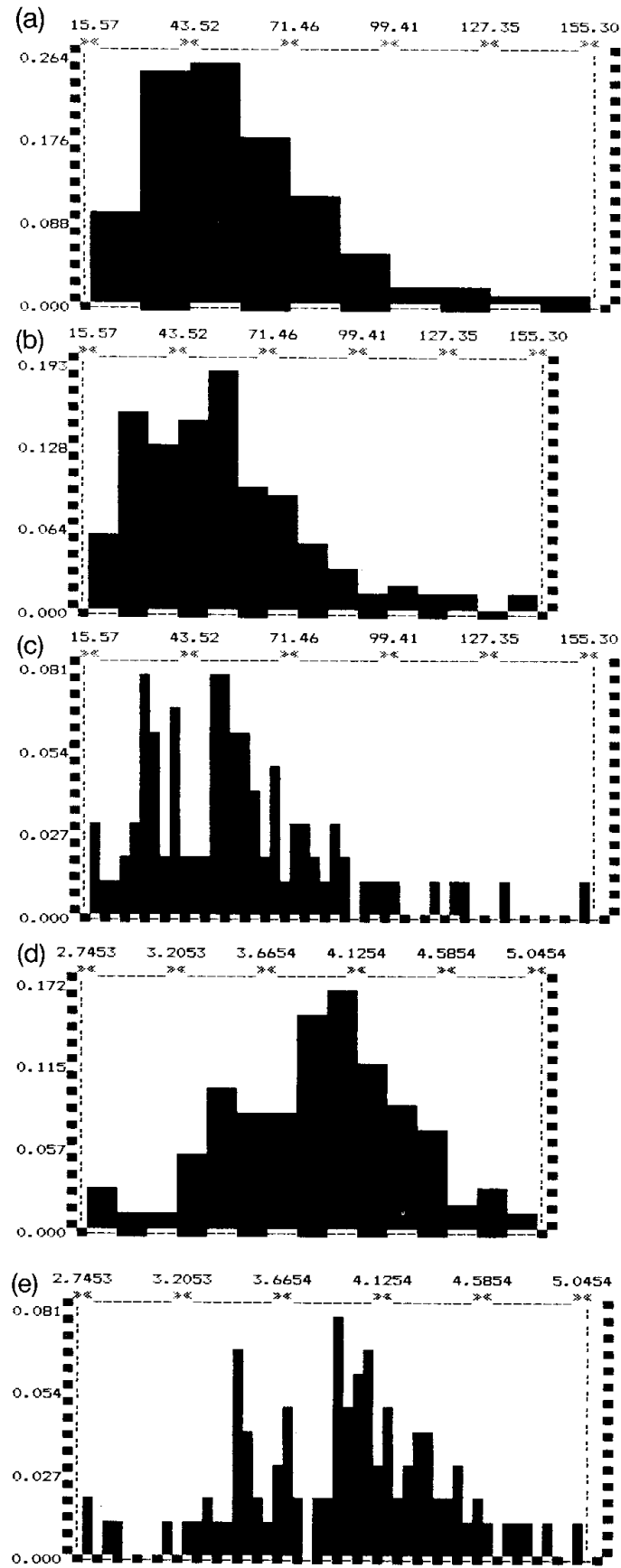


圖 7-1 a~c 為改變等級數目的變化；d~e 為轉換方式改變的變化

### Exogenous versus idiographic class intervals :

一個非常廣泛用來比較 exogenous class limits 的成功的方法，是把資料分配至各等級內，然後用 one-way 變方分析來估計，用等級(classes)來解釋全部資料組的變方有多少？

model:  $Z(X) = \mu + X_j + E$

E: 在 class<sub>j</sub> 裏的點 X 的可變值

$\mu$ : 全部平均值

X<sub>j</sub>:  $\mu$  與 class<sub>j</sub> 平均間的變異

E: 常態分布 error 項，平均為 0，變方為  $\delta^2$

好的分類: class 平均之間的差，越大越好。剩餘機差 E 越小越好。

常假設每個 class 間的組內殘差(residual within class variance)相同，表 7-1。一個簡單方法來估算分類的效果，就是計算  $S^2_w/S^2_t$  比值，稱為相對變方。 $1 - (S^2_w/S^2_t)$  估計變方的比率(分類之說明)其值越小越好。

Exogenous 分類的成功大部分依賴採樣區的資料，及介於性質種類之間及使用的區別規則二者間的關連性。

### Visual appearance and information content in choropleth maps :

地誌圖是利用一些水平平面切割資料而可看到的結果，其位置由邊界分級而設定。最好是(可能的話)把原始資料輸入 GIS，或者至少捨棄尚未被可信賴的知識(關於組內平均變異，within-class means and derivations)所支持的分類好的資料來源。

### 多變量分析及分類：

#### 排序(Ordination)：

為便於了解，把複雜多變異的資料減量是有必要的，主成分分析是最普遍應用的方法，來檢查資料以比率比例尺則定以便交互關係。

主成分分析：以數學技巧來檢查 n 個別項(每項各有 m 個性質)之間的關係。

$$PC1 = \alpha_1 V_1 + \beta_1 V_2 + \gamma_1 V_3 + \dots + w_1 V_m$$

$$PC2 = \alpha_2 V_1 + \beta_2 V_2 + \gamma_2 V_3 + \dots + w_2 V_m$$

$$\vdots \quad \quad \quad \vdots$$

$$PCm = \dots\dots\dots$$

把資料利用旋轉、投影至新軸上，常發生在不同成分可以物理內插主成分的名次可以當成新變數，在前 6 個主成分中常包括了 20-40 原始變數(有時可達 70-80%)，可非常有用的描述資料組成。

### 集群分析(Cluster analysis)：

Ordination 的主成分分析，雖把資料簡化，但卻是轉換成為一種形式，而 cluster analysis 則是把資料簡化外，在二種座標軸都有近似的值。今日，cluster analysis 是項標準技術，可用於多數大電腦，通常要把資料由 GIS 送到 clustering programs 並不困難。

開始進行：

divisive system：全部資料組檢查，以最好方式細分成幾群。

Agglomerative system：在把各別資料點合成一群之前，檢查其相似性，點的距離越接近，其相似性越大。

第一步：similarity 的檢查：

the Euclidean distance： $d_{ij} = \sqrt{\sum_{k=1}^m (X_{ik} - X_{jk})^2}$

X：K 性質的值在點 i、j

缺點：對變數測定的比例尺很敏感，因此通常在計算資料時，先由其同樣關矩陣轉換至主成分計算，主成分是平均為 0，變方為  $\delta^2$ 。

The Canberra metric： $d_{ij} = \frac{1}{m} \sum_{k=1}^m \left[ \frac{|X_{ik} - X_{jk}|}{X_{ik} + X_{jk}} \right]$

範例 0-1 為 dissimilarity 相似為 1-d<sub>ij</sub>

the city block metric： $d_{ij} = \sum_{k=1}^m |X_{ik} - X_{jk}|$

Mahalanobis's  $D^2$ ： $D_{ij} = (X_i - X_j) \Sigma^{-1} (X_i - X_j)$

$X_i, X_j$ ：點 i, j 名次的向量

$\Sigma$ ：pooled within-groups variance-covariance matrix

優點：接受變數間之其同相關

其他方法在獨立變數應用較佳

這些方法所建的矩陣大小是  $1/2N(N-1)$  N：資料組的 item 數

1000 筆 items 需要 2M byte for 32bit

第二步：把個別資料連接至 group 上

the single linkage method：點分佈至其最接近的鄰居所屬的 group

the centroid method：item 結合至距其最近之 group center 的 group，每個 item 加入後，group centroid 要再重新計算。

Ward's method：項目(item)加入於群組(group)內個別值與 group 平均 derivation 平方和增加值最小的 group。

在計算相似性(similarities)及建立項目(items)及群組(groups)間連結(linkages)方法的範圍導致大量可能分類策略情形。理想上，如果資料結構良好，不同方法也會產生類似結果，實際上則不然。

Ward's method 可用於概念清楚，定義良好的 cluster。

Single linkage method 則用於 cluster 幾年不存在的概念。下一個問題是 cluster 分析時用來總結結果的 cluster 的數目，Webster(19 )用 Wilks's criterion 統計來解決。

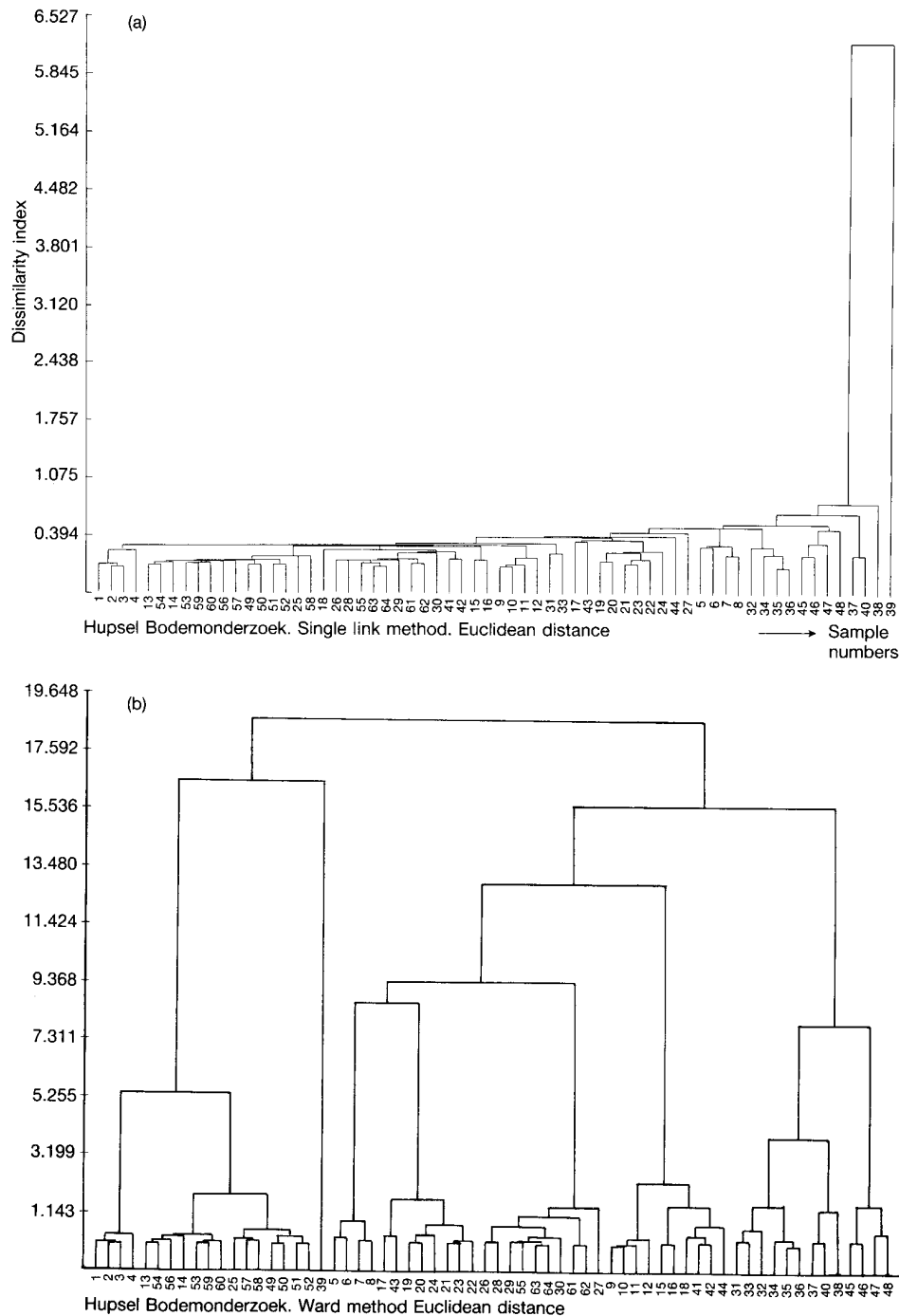


圖 7-2 (a)簡單連結法(b)Ward's 方法

我們希望這些分類方法可用在地圖上顯示不同等級數的位置。Webster and Burrough (1972 a, b)嘗試用 cluster 分析的結果用在土壤資料，但結果不佳。圖形太差點狀化，干擾太多。

De Gruijter (1977)把原始資料內插入較小網格，然及用分配(allocation)方法已分級之成立原始樣品資料來逼近網格單元(match grid cells)，結果比前述 W and B 佳，但仍不理想。

UK's Irstitnte of terrestrial ecology：同時用 cluster method 及 allocation method

生產小尺度鳥瞰及生長地圖來計劃研究，並沒有產生像土壤研究之點狀的問題。

最後一個問題是 clustering methods 是特別記號的，其結果依賴資料組正確的性質，同時分類導出之來自鄰近區域的資料並不能正確的比較，即使 clustering 策略是相同的，由這些分析得到的 cluster 之物理項要內插是很困難的，同時要例行調查也是不被喜歡的。

Allocating individuals to existing classes :

GIS 多數使用者並不是要建立新分類，而是把資料分別放入已存之分類中，這個問題很容易分析，因為通常有一些屬性定義為變值空間，對於偏差之討論並無事先存在之分類。

最大概似法(Maximum likelihood method)是本節討論方法中最有差別的，其缺點是比其他方法需要更多運算。

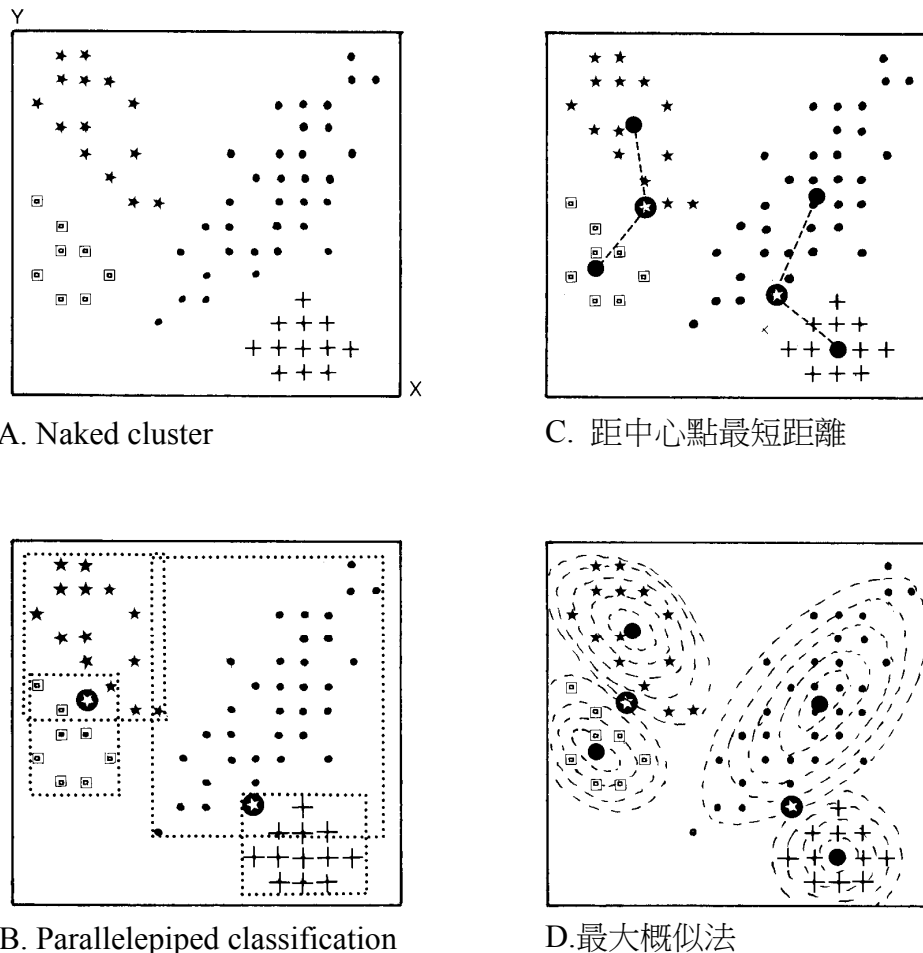


圖 7-3 不同的集群方法會有不同的結果

GIS 之專家系統：

分類及分配過程之應用係確使使用者及複雜資料庫中擷取 item，如果 item 輸入前已經被分類，那麼這項工作就不重要了，使用者又要分類，又要分配資料，這並不容易，因此借助電腦系統來完成這項工作，稱為專家系統。

原理是專家把其知識使用至電腦，稱為 knowledge acquisition 所需的知識轉

換成一些規則存在於知識庫(knowledge base)中，供系統讀取。

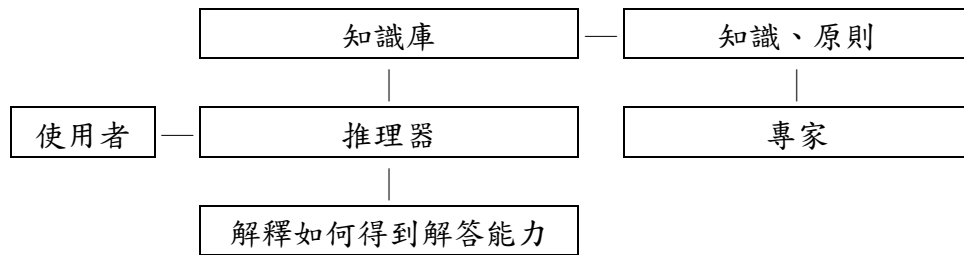


圖 7-4 專家系統概念圖

最初，知識是以正式邏輯敘述的形式 coded，人類用更多修飾過的方法。如類推，現代建立人工智慧進行系統構能捕捉更多人類思考過程觀念。

使用者由 knowledge base 得到知識，由程式中達到其需求稱為 inference engine，這個模式把使用者需要轉換成公式化需求可被需要的知識使用進行 knowledge base，inference engine 有解釋設備，可告訴使用者為何它正在尋找特別種類的實體。

專家系統應用於醫藥、地理探測、軍事，將利用至土地評估，基本上專家系統可利用知識庫，由世界級或洲級真實專家基礎，所以地區擴展及建立工作者可以由最佳設備獲查。

Summary-GIS 分類方法：

分類方法是任何 GIS 必要包括的部分，許多現代的系統只提供了外部的、隨意的及等間距的分類，為了科學研究影像分析系統，將有多變數分類方法以樣式及外貌察覺或增加數值分類多變數方法專家系統，為未來提供了希望。

分類的規則：

1. 不分類：(a)如果你有一塊面積完整的空間圖層，以及原始資料值在進行過程中會被使用。(b)如果在樣品點收集了資料，先用內插提供資料到全部面積，如果必要的話就分類。

2. 分類：如果資料是多變數，且如果用一個物體或形式做 cluste 鑑定是必須的，鑑識對使用者是有用的，但是在現在資料庫外部。

3. 分類：為了展示及描述空間資料，但不是為了資料儲存。不要使用任意比例尺然後謹慎地對待比例外的東西，特殊比例尺度有時是有用的，盡可能用客觀定義的連續比例尺。

## 第七章 分類方法

地理資訊處理過程中的資料分析可說是將空間資料轉化為我們所能瞭解的資料形式，通常稱為結構（structure）或訊號（signal）。而我們無法描述的稱之為雜訊（noise）。而主要的問題在於，這些結構或訊號是我們所能辨識、分級及套疊的，亦是資料萃取及數量化分析的不變的傳統調查方法。

在傳統的調查與套疊時，原始資料通常可以很快的將之分類。最常見的為土壤調查。土壤剖面經描述與分析後套入適當的分類標準系統中，地形一致及剖面分類相同者為同一單元。而使用航空照片時，處理過程有些不同，即在地面調查之前就先繪出土地單元輪廓圖，雖然如此，最後的結果亦大致相同。土地分級是參考土壤分類中的分級單元來決定其為何等級。相似的方法已被應用在整合性調查與其他自然資源調查。

當人們沒有很有效的方法時，面對大量且複雜的資料時要快速的簡化資料上述的過程是不可避免的。確實，在這種情況下大概只能用感覺的。方法論中假設透過對一物件的研究瞭解（如土型、土壤質地等）便可將自然單元加以確認，並將其一般化以利資料的儲存與轉換。這個方法的成功與否在於分類中的關鍵：爭議與可分辨的屬性之間的關係及使用者真正需要的大量資訊。如果這兩者間的關係薄弱或不存在，那這個分類方法將幫助不大。

許多地理資訊系統呈現以刻度描繪地形輪廓的資料結構時，主要的資料來源為傳統的主題圖。這個刻度描述整個模型或疊圖時分類項目為依據的邊界所產生爭議點的平均值。在刻度或研究報告中，空間變異及套疊時屬性值變異只可能被暗示。原始資料會全部被形式面積說明（type area description）或剖面模型（modal profile）及其他標本（representative）實體所取代。結果導致原始資料（所有空間資料的結構與尺度）被人工的線條所取代而成為只有一種結構或一種尺度。因此在圖內所有精確的數量資料的變異都會漏失。

雖然這些種類的主題圖能夠表現大量面積的空間資料變異，然而卻只能處理地區性，圖內的單元在自然疊合處理過程中已消失。在人工套疊處理過程中，取決於尺度、結構及屬性。然而，在地理資訊系統中，我們不需要在資料尚未分析來使用之前就將之分類或簡化。相反的，我們可以直接在 GIS 中將之排序，因此我們可以就特定目的予以最佳的分類或使用適當的內插法來套疊單獨的屬性。因此 GIS 可以選擇資料庫中控制的子集合並選擇最佳組合及內插法。

下兩章的目標在於描述使用者從原始地理資料中利用適當的統計方法及數量化方法萃取出更多的資訊的方法。如此做會使 GIS 所進行的分析與模式結果的邊界誤差情形改善。



## 分類 (classification)

分類是讓人類瞭解的基本方法，沒有分類將無法對細節更瞭解。人類頭腦對分級亦有其接受程度，一般為 7 級上下兩級。而以二分法而言，又不夠敏感，十個等級以上的分類又過於複雜無法運作，因此等級通常為 7~10 個等級。

### 單一連續變數值的範圍分類

簡單的分類方法常用於顯示變動的數量化空間屬性，如人口密度、土型坡度、空氣污染等級等易於瞭解的主題圖。原圖的等級區間的選擇便非常重要。Evans (1977) 對於他自己的分類系統中的分類區間有深入的研究：

**Exogenous class intervals**：根據有關而不是分離的資料組中起始值開始，等級區間是固定。雖然其為標準分類系統中的一部份，卻不需要是問題中最適當的面積。Exogenous class interval 廣泛應用在土壤調查及土地評價分類系統。最大優點在於應用廣泛，缺點為在一特定地區無法解決變異。如以土地型態坡度分級為例，0-2%，3-5%，6-8%，9-16%，17-30%，>30%，以足以簡單的將土地型態坡度分級。

**Arbitrary class intervals**：無特定目的時採用，這種區間常是不規則的配置且資料也未經適當檢驗。

**Idiographic class intervals**：考慮資料組特定目的的關係時採用。包含將資料分成多模型群組的方法。

**Serial class intervals**：兩值間極限的直接數學關係。如果這些極限值決定於特定資料組的屬性，其結果是相似的。變數圖可將之想像為一統計表面，如下敘述：

- (a) Normal percentiles：等級區間再細分為一常態分佈 ( $\mu, \sigma$ )。
- (b) 以標準差比率來區分等級中心點為平均  $\mu$ ，如  $\mu - 1.8\sigma$ ， $\mu - 1.2\sigma$ ， $\mu - 0.6\sigma$ ， $\mu$ ， $\mu + 0.6\sigma$ ， $\mu + 1.2\sigma$ ， $\mu + 1.8\sigma$ 。這種分級區間可以應用在偏歪的資料或包含真實多模型的分佈或一些離異值。
- (c) Equal arithmetic intervals：區間寬度相同。其主要的問題在於等級區間可能為十進位且難以比較，因為每一資料組都有特定屬性，來自不同資料組的同樣變數的地圖將會有極大的不同。
- (d) 倒數、三角函數、幾何或對數尺度的等間距：當原始資料為非常態或矩形分配時可使用此法。

圖 7.1 改變等級的數目 (a-c) 或改變轉換方法 (d-e) 可看出不同的處理對一資料組的影響可能成為另一種形式亦可能更自然的分級

## Exogenous versus idiographic class intervals

問題：Exogenous 範圍較佳或將資料維持某些自然結構較好？

常用來比較 exogenous class limits 的成功與否的方法，通常是用這兩者來分割資料，然後作 one-way 變方分析來估計總資料組能解釋的變異量分析模型如下：

$$Z(x) = \mu + \alpha_j + \varepsilon$$

$Z(x)$  為變數  $Z$  在已知  $j$  等級中  $x$  點的值，由平均  $\mu$  及  $\mu$  與等級  $j$  的平均  $\alpha_j$ ，及殘差  $\varepsilon$  構成。

這個模型的假設為在資料頻度分佈的每個模式能在此等級中成立。

好的分類：等級間的平均相差越多、殘差越小。而我們通常假設每個等級中的殘差都相等，如表 7.1。

表 7.1 one-way 變方分析

有一簡單的方法可以估測分類的有效性，即  $S_w^2 / S_t^2$  比值，即所謂的相對變異。

$S_w^2$  為組內變異， $S_t^2$  分類前母數的變異

因此我們可以下式估計分類時的變異部份：

$$1 - S_w^2 / S_t^2$$

Exogenous 定義的分級可透過相對變異予以比較，而已知目的所回應的最低值的那一個是最好的。Exogenous 分類成功與否取決於資料、由地面取得的樣本及屬性種類與可辨識準則的關係。

## 地誌圖中的可見外觀與內文訊息

地誌圖是利用許多水平平面切割資料組所得的結果，其位置由邊界分級所決定。地理資訊系統中使用的資料來源為許多主題圖，這些主題圖並非只是資料分析與分類。這些主題圖資訊可透過人工巧妙的處理而輕易得到，但是事前分類若不夠周延是一件危險的事。因此最好是盡可能輸入原始資料或盡可能將無法提供分級的資料捨棄。

## 多變量分析與分類

在許多環境調查中我們所面對的是一組資料點的許多屬性值，因此必須透過簡化來更瞭解。主成分分析常用在處理比例尺度以瞭解其間的關係。另外集群分析將多變量資料組中明顯易見的群組集結或將兩軸相似資料點集結。因此而減少複雜的資料組進而將其集群以作為分類的準則。因為集群分析是使用原始資料，因此，可將資料的原始結構顯現出來而不失真。

在做前述分析時第一步先將資料組細分後，再選擇最好的方法予以分群。Agglomerative system 可將兩個個別變數找出相似性，並將其歸於同一群。而 M 度空間中一對資料的相似性有許多方法可以估計，如以變數間的幾何距離。如下式表示：

$$d_{ij} = \sqrt{\sum_{K=1}^M (X_{iK} - X_{jK})^2}$$

X：k 性質在 i、j 點的值

缺點：對變數測定所使用的尺度敏感，因此通常由相關矩陣所得之主成分來轉換計算此一距離。另一點，此一距離對極端值或離異值亦敏感。

Canberra metric：

$$d_{ij} = \frac{1}{m} \sum_{k=1}^m \frac{|X_{ik} - X_{jk}|}{X_{ik} + X_{jk}}$$

0-1 之間為不同的值，1-dij 為相似值

$$\text{City Block Metric} : d_{ij} = \sum_{k=1}^m |X_{ik} - X_{jk}|$$

$$\text{Mahalanobis's } D^2 : D_{ij} = (X_i - X_j)' \Sigma^{-1} (X_i - X_j)$$

$X_i$ 、 $X_j$  分別為在 i、j 點的分數向量； $\Sigma$  為組內變異數-共變數矩陣

其中 Mahalanobis's  $D^2$  優點為允許變數間存在相關，而其他方法則必須為獨立變數。

接著，相似矩陣算出後，即可將個別資料連結到群組中。最明顯的就是找到點與點在空間定義上接近者，這很容易。可是要在找第三個或更多來加入已有的群組中便很難。

- 1.簡單連結方法：可找尋最鄰近的點或群落來加以連結。
- 2.重心法：以此點距離某一群組之中心最近者，然若在加入一個項目，則所有距離都要重算。
- 3.Ward's method：加入一項目時，使群組間的變方和的增加最小。

以上的方法可經由樹狀圖或連結樹圖來連結。

理論上而言，資料結構越好，不同的方法所得的結果相去不遠。圖 7.2 說明應用集群方法處理 64 種土壤剖面的兩種樹狀圖，每種剖面有 20 個項目，資料經過主成分分數的轉換，並計算其幾何距離，只是連結的方式不同。兩種結果並沒有差很多，可是 Ward's method 所得之圖可達到清楚、定義較佳的集群，而簡單連結法卻無法將集群明顯分出。

圖 7.2 (a) 應用簡單連結法所得之結果

圖 7.2 (b) 應用 Ward's method 所得之結果

接下來討論當使用集群分析時，要決定到底要有多少個群組。Webster(1977)建議可以使用 Wilk's criterion 來達到此目的。

另外我們也希望分類方法應該達到使一張地圖顯示不同分級的組成顯示出其位置。Webster and Burroughs (1972a) 嘗試應用集群分析於土壤資料，可惜結果並不佳，即使加入空間資料組或區域空間加強也改善不多。De Grujter (1977) 採用不同的方法，首先將原始資料以內差法求得較佳的網格，接著以一從原始資料定位的網格資料予以套合，其結果比 Webster and Burroughs 所得較佳。另一方面，UK's institute of Terrestrial Ecology 採用集群技巧利用定位的集結方法來產生小尺度的地景圖及植生植栽計畫研究等。

集群分析的最後一個問題是其有個人使用特性，從資料組所萃取出的屬性都是獨立的。大多時候，由原始資料分類都以鄰近的區域而不是經由兩兩比較而得，集群分析亦是。因此要解釋這些分析所得到的集群的物理項目是很困難的，而在例行調查的分析卻有人喜歡採用。

### 將個別資料定位於存在分級中

對許多 GIS 使用者而言，問題並不在於建立新的分級，而是在於如何將個別資料適當的定位在已存在的分級系統中。也就是許多等級被定義多於一個屬性的值的偏差，甚至等級無法套疊，也就是說在多變量等級套疊及合併時會有灰色

地帶產生。

欲解決此一問題，必須先對屬性作科學分類（可參考植物分類或動物分類）或採用許多國家中的土壤分類形式的方法（在土壤調查中特別如此）。為特定目的所採用的分類的標準方法似乎應該在為另一目的時直接了當的切入其需要，因此使用者可以更滿意其結果。

在衛星影像分析時個別資料定位的問題很容易分析，因為只有少數變數為多變量（如 4，7，或 16 波段）且並沒有以存在的分類偏誤的討論。如圖 7.3 (a) 顯示 X 和 Y 定義在兩度空間中的四個集群點，至少有三種方法可以將一新的個別資料定位於存在的分級中。圖 7.3 (b) 由 X 和 Y 屬性的極端值所定義的邊界等級。此即影像分析中所謂的平行六面體（parallelepiped）分類，這跟採用簡單布林邏輯分類相同的原則。因為集群不是球形但卻是拉長形，因此套疊時白點與黑點無法很明顯的表現。圖 7.3 (c) 採用另一種方法定位原則是根據該點離集群中心之最短距離。圖 7.3 (d) 採用每個族群計算過機率的鐘形區域，選擇關係最大者，如最大概似法。

圖 7.3 不同分類方法對多變項空間分佈形狀所產生的影響 (a) 無根據的集群 (b) 平行六面體分類法-白點和黑點分佈分歧 (c) 與中心距離-依然不明顯 (d) 最大概似法-為等機率線

採用 Bayesian algebra 根據每一等級的優點機率加以權重可延伸最大概似法。權重的使用可將 misclassification 的成本最小，而達到最佳分類。最大概似法在此為最可分辨的，其缺點在於需要比較簡單的方法更多計算。

## GIS 的專家系統

分類的目的與定位過程是為了讓使用者能糾正複雜的資料庫。如果資料在輸入之前就已經分級，那此一任務就很平常。進一步想，大量未分級的資料輸入系統中，使用者便必須對分類與定位點予以定義來找出什麼是必須的。如上所述，這些過程並非簡單的事，結果也並不是百分百令人滿意，且一個臨時的系統使用者或非專家在工作的一個特定區域便不能有系統的找出其需求。能做到上述工作的專家並不多，因此如何將專家的知識轉化輸入電腦系統中以便讓非專家的使用者使用的系統建構即為專家系統（expert system）。圖 7.4 顯示專家系統的一個簡單的基本概念。專家系統的法則即為將具有某方面知識且知道問題所在的專家所具備的知識輸入電腦中，這些知識必須經過辨識或描述特定實體的規則化後電腦始能接受。剛開始時，知識以一般邏輯敘述輸入電腦中，而人類當其面臨問題時會有許多種解決的方法，這便可以用一推理器來達到。近年來人工智慧朝向能擷取更多人類觀念的系統發展。

圖 7.4 專家系統架構

使用者欲從資料庫中獲得資料，只要透過推理器的運作便可達到。這個模組具有將使用者需求系統化，而推理器具有解釋能力，可以告訴使用者為何得到此一解答。

專家系統廣泛應用於醫藥、生化、軍事等方面，因此將專家系統應用於土壤分類，將專家的經驗延伸，透過邏輯的描述，將複雜的因子予以綜合考慮，並將之規則化，以建立國內的標準。然而，專家系統往往與限於地區性的使用。

### 總結-GIS 的分類方法

在任何 GIS 系統中，分類方法佔有重要地位。然而現在有許多的系統多採用布林邏輯，並指提供 exogenous、arbitrary 及 equal interval classification。影像分析系統在模型辨識具備的多變項分類法及未來科學研究採用 GIS 時可望能被涵蓋進去。專家系統在未來將有前景。

#### 分類的幾項原則建議如下：

1. (a) 如有一地區的完整空間資料及原始資料在進一步處理時會用到時並不需要將其分類。(b) 如所蒐集到的資料為一組簡單的點時，則先採用內差法求出整個區域，之後若有必要再予以分類。
2. 如果資料為多變項或是以一個物件或形式確認的集群再進一步分析很重要時，必須予以分類。確認工作可讓使用者得到所需資料庫以外的資料。
3. 將要展示的空間資料分級，而資料儲存時並不需要分級。不可用不能變動的尺度且在採用外來尺度時要謹慎。有時候獨特的尺度也許有用，但應盡可能使用序列尺度所定義者。

## 第八章 空間內插的方法

Interpolation：在已觀測到點範圍內未採樣位址性質的估計過程稱之。

Extrapolation：在已觀測到點範圍外未採樣位址性質的估計過程稱之。

前提假設：空間中點的位置越接近，其性質也越相似。

本章在描述一些最普遍常用的內插方法，並推測一些方法，這些內插技術可幫助當嘗試把點及面積資料帶到空間網路進行地圖疊合及製圖模式化。

### 內插可用的方法：

內插的問題是選擇一個可靠的模式來模擬資料的問題。

#### 藉由畫邊界內插：

最簡單的內插方法是用地貌外特徵繪出地景單元(landscape units)的輪廓，用在多數地誌圖、土壤、地質、植生的主題圖、類似技術是 edge-seeking algorithm 用於影像分析，這些技術的前提是所有重要的變異都出現在邊界，在邊界內其變化是均一且具同向性的，導致階層性的模型。

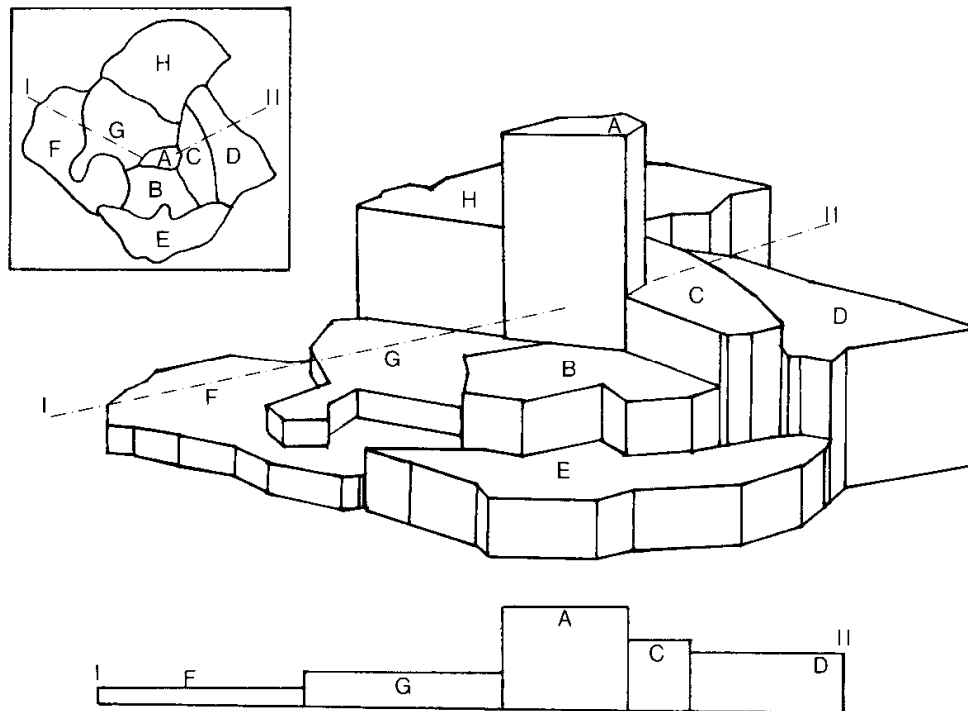


圖 8-1 主題圖為一階層模型的地景

另一個內插的方法是用 Thiessen polygons，又稱為 Voronoi polygons or Dirichle cells，這是二維的，未知點的資料取用最靠近它的已知點資料，如氣象資料取用最近氣象站的資料。

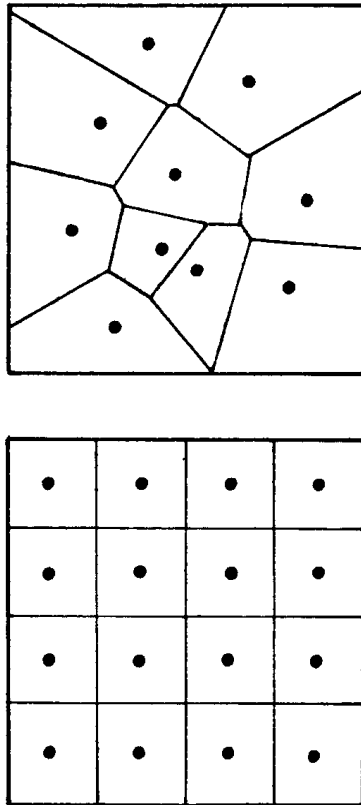


圖 8-2 以規則與不規則的 Thiessen polygon 來作取樣點的研究

Green and Sibson (1978)提供數學方法來計算這些 cell，Tipley(1981)討論這些方法而有以下缺點：

- 1.面積大小及形狀靠樣品排列而定，會造成 polygon 有奇怪的形狀，尤其是在邊緣；
- 2.每個 cell 的性質由單一樣品估計，無法了解估計的誤差邊際；
- 3.未採樣點值的計算成為一點 point-in-polygon 的問題，與線間的想法-越近的点，性質越接近的想法無關。Sibson(1980)以加權平均法建立 natural neighborhood 內插。

逐漸變化的內插方法：

包括 Spline Functions, Least Square Trend Surfaces, Fourier Series，移動平均包括 kriging，這些方法可分為全球性及區域 fitting 技術。

Global：模式建立在研究面積內全部觀測點，如 Trend Surface Analysis 及 fourier Series 用於長範圍變異如土地形態地下水位。

Local Fitting：splines, moving averages estimated value 由鄰點，local 反常值不會影響表面上其他點內插值。

Global methods of interpolation：

1.trend surface analysis



描述漸進、大範圍變化的最簡單方式就是用多項式迴歸來模型化，

迴歸線被  $\text{fit} \sum_{i=1}^n (\hat{Z}_i - Z_i)^2$  最小化

二維時， $f(X,Y) = \sum_{r+s \leq p} (b_{rs} \times X^r \times Y^s)$

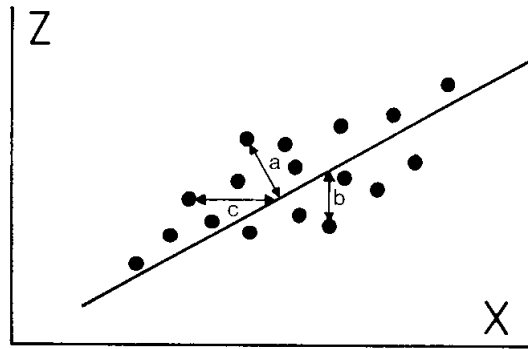


圖 8-3 趨勢分析中，迴歸線以距離 b 最短來形成曲線

P 維時， $\sum_{i=1}^n \{Z(X_i) - f(X_i)\}^2$   $X_i$ : 向量 for(X,Y)

*flat*  $b_0$

*linear*  $b_0 + b_1X + b_2Y$

*quadratic*  $b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2$

找  $b_i$  係數是多次迴歸的標準問題，很容易進行。

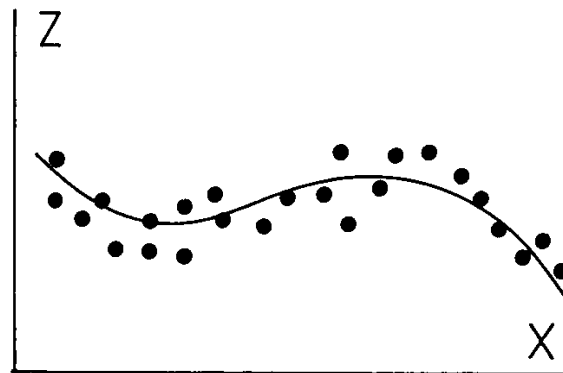


圖 8-4 多項式在點的線性分布逼近時是必要的

由估計規則網格上每一點  $Z(X)$  的值，可展示 trend surface。

Trend surface analysis 的優點：很容易了解，資料的許多性徵可由低次 trend surface 模擬。Trend surface 是平滑的函數。

Trend surface 的主要用處：並不是在區域內內插，而是在進行其他 local 內插之前移除大量資料特徵(feature)。

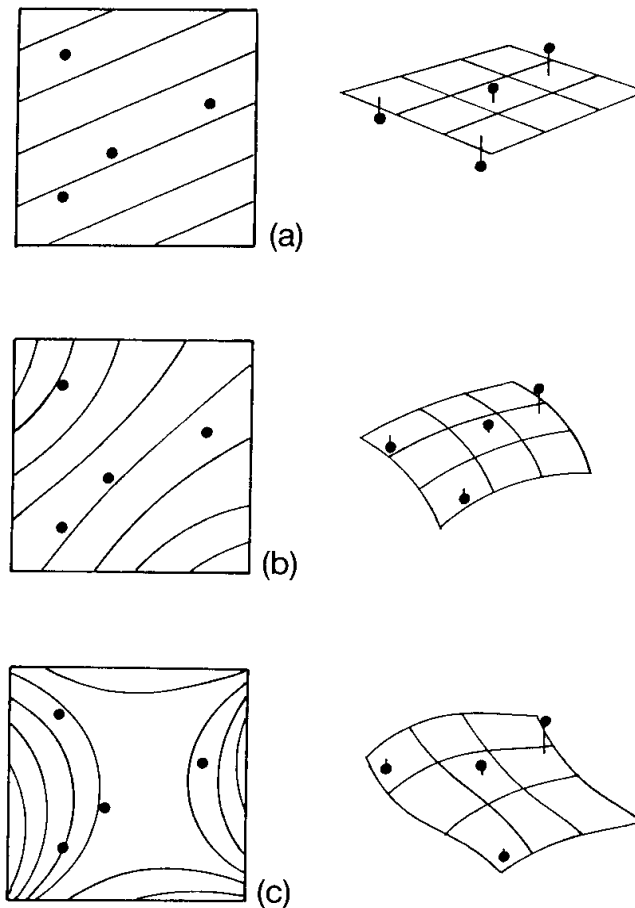


圖 8-5 在二空間中的趨勢面：(a)線性(b)二項式(c)立體

趨勢面的顯著性(The significance of a trend surface)：

進行變方分析 F test：表 8-1

Source	SS	df	MS	F
Regression	SSr	M	MSr	MSr/MSd
Derivation	SSd	n-m-1	MSd	
Total	SSt	n-1	MSt	

Regression relative variance  $RVr = MSd/MSt \times 100$

相對變方可用來比較地誌圖與 trend surface。

2. Fourier series：

由 sin 及 cos 曲線組合把觀測變異模式化而描述一維或二維變化，一維應用於時間來到如氣象變遷，二維則用於沉積地質 Fourier series 用於結構分析多於地圖，Fourier series 不適用於土壤調查內插。

地域性推估法(Local interpolator)：

1. Splines：

在電腦用於配適曲線前，作圖者用彈性曲尺目視來 fit 平滑線，這種彈性曲尺稱為 splines。

Spline functions 可正確地 fit 少量資料點，同時卻使曲線為連續性，亦即用 splines 可以修正一部分的曲線，而不必重新計算全部，

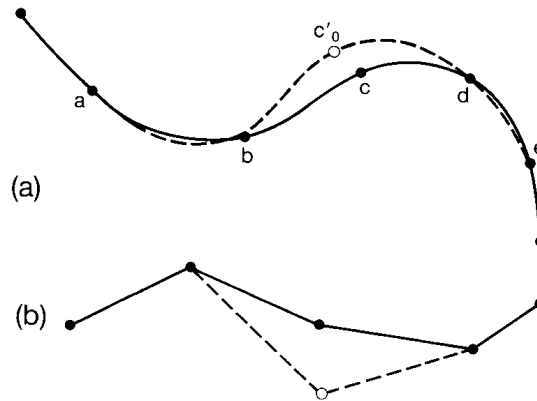


圖 8-6 當其中一點改變時，二項式(a)必須重新計算四個區段，曲線尺規(b)只需重新計算二個區段

這在 trend surface on Fourier series 是不可能的。

定義：

$$P(X) = P_i(X) \quad X_i < X < X_{i+1} \quad i = 0, 1, \dots, k-1$$

$$P_i^{(j)}(X_i) = P_{i+1}^{(j)}(X_i) \quad j = 0, 1, \dots, r-1 \quad i = 1, 2, \dots, k-1$$

點  $X_1, X_2, \dots, X_{k-1}$  把區間  $X_0, X_k$  分成  $k$  段，稱為 break points，曲線上這些點稱為 knots。

B splines 常用來將展示數化線平滑化，如土壤及地理圖邊界。使用 B-splines 在平滑多邊形邊界上，會造成一些爭議，尤其是在計算面積及周長，如果多邊形面積是由數化資料點利用梯形定理計算，會與由 B-splines 平滑邊界所得的面積不同，另一個問題是多項式在平滑彎曲邊界包含了尖的、矩形角。

用 splines 內插時有一個問題，就是要選擇轉折點(break points)去符合資料點或插入它們之間。

Splines 用於正確內插(splines 函數穿過所有資料點)或用於平滑化，後者是一個敏感的過程，當資料點的值有實驗誤差時。

$$Dubrule(1983, 1984) \quad Y(X_i) = Z(X_i) + \sum (X_i) \rightarrow \text{隨機機差}$$

$$A(f) + \sum_{i=1}^n W_i^2 [f(X_i) - Y(X_i)]^2$$

$Z(X_i)$ ：點  $X$  的屬性測值

$A$ ：函數  $f$  的平滑線

第二項：資料精確度

權重  $W_i^2$  以機差反比表示

$$\text{權重} \quad W_i^2 = \frac{P}{\text{Var}[\sum (X_i)]} = \frac{P}{S_i^2}$$

$p$  反應 splines 曲線重要特徵

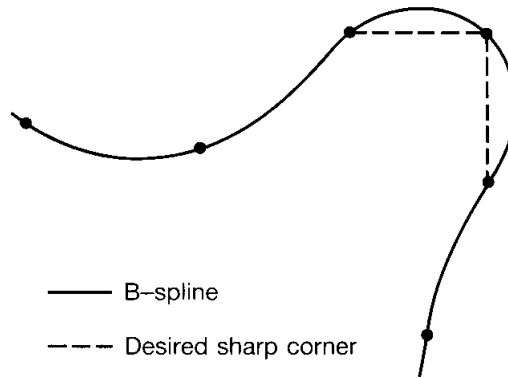


圖 8-7 曲線尺規法使用在邊界時是不方便的，會出現曲線和角狀的情形，除非使用指示點

*splines* 的優缺點：

優點：內插計算很快，特徵是小尺度，快速產生清晰的資料總觀

缺點：*splines* 內插時無法直接估算誤差

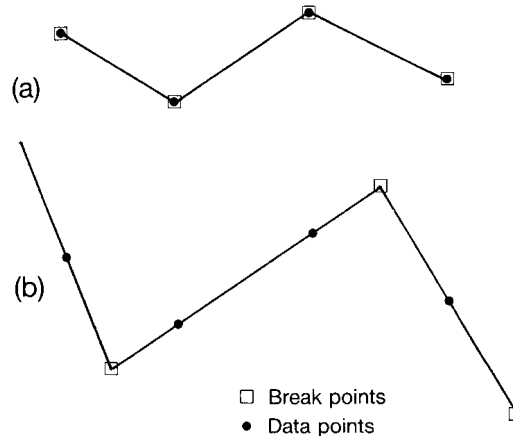


圖 8-8 曲線尺規法的形式通常取決於轉折點(a)轉折點在資料點上(b)轉折點不在資料點上

用移動平均法：

在內插一個變數  $Z$  於未知點  $X$  的常用方法之一，是由鄰近範圍為或稱 "Window" 計算平均值，最簡單的形式：

二維時  $X_i$  由  $X_i$  取代

$$\hat{Z}(X) = \frac{1}{n} \sum_{i=1}^n Z(X_i)$$

$\hat{Z}(X)$ ：點  $X$  在對稱 window 中央的 moving average

window 大小對平滑的輸出形式有絕對影響，窄的 windows 造成小範圍變異，寬的 windows 會增加長範圍效果。

加權 moving average：相距近的觀測值比距離遠的觀測值要相似

$$\hat{Z}(X) = \sum_{i=1}^n \lambda_i Z(X_i) \quad \sum \lambda_i = 1 \quad \text{權重 } \lambda_i \propto \phi(d(X, X_i))$$

當  $d \rightarrow 0$  時， $\phi(d) \rightarrow \infty$  出現在倒數或指數函數

加權平均內插造成的表面依賴所用的函數或其參數以及樣品資料點所繪定義域或 window 的大小。

定義域的大小不是影響點估計的平均值，還控制了內插所需的電腦時間。

不規則資料造成團塊叢生的問題，可由降低那些資料權重方式來修正。

在格點內插值可直接以灰階(grey-scale)網格地圖或彩色網格螢幕或印表機展示，內插 grid cell 值可用於網格資料做為圖層。

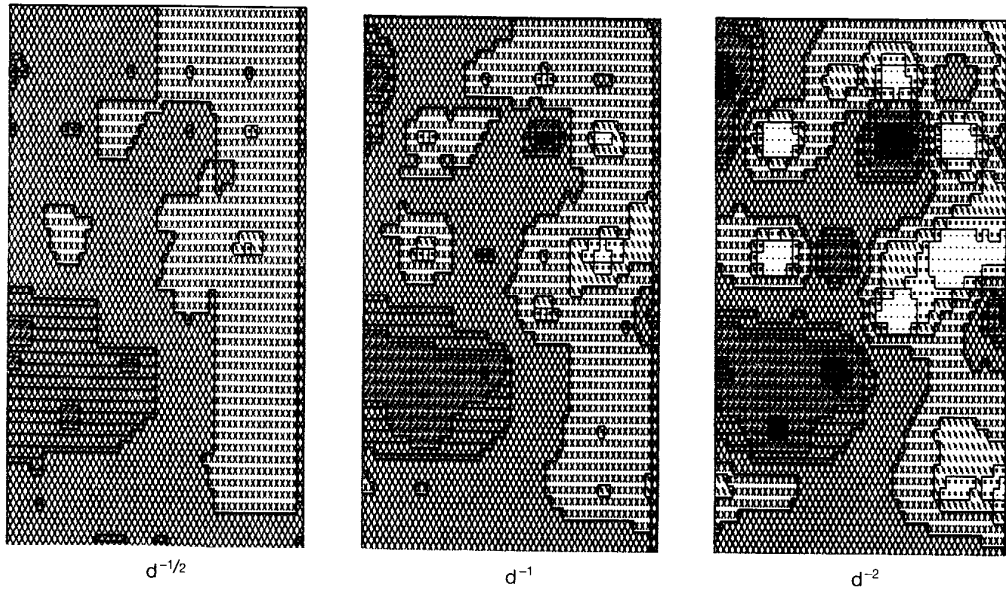


圖 8-9  $r$  值不同距離權重即不同( $d^{-r}$ )

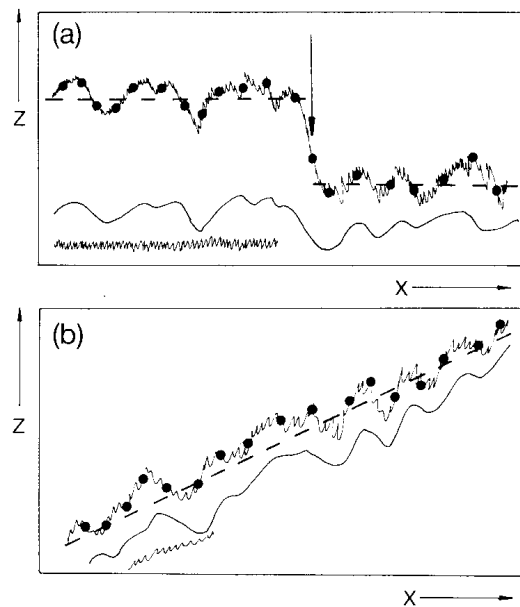


圖 8-10 空間變異的主要成份

-----結構值  $\sim \sim \sim$  空間相關隨機差  $\varepsilon'(x)$

$\sim \sim \sim$  非空間相關隨機差  $\varepsilon''(x)$  (雜訊) · 取樣點(a)變化劇烈(b)穩定

利用空間自動變方視覺內插方法：

local weighted moving average interpolation methods 產生下列重要的不確定性：

- (a) domain or window 要多大？
- (b) 適當的內插(sample sites)應是什麼形狀及位相？
- (c) 估算  $\lambda_i$  權重會比簡單距離函數好嗎？
- (d) 隨內插值而來的誤差(不確定性)是什麼？

BLUE：最佳線性無偏估值

方法留下出"區域變化"太不規則變化無法以平滑的數學方程式模式化，可以猜測的表面來描述，結果是用來估計內插權重  $\lambda_i$ 。

區域變化的理論假設任何變化的空間變異，可以三個主要成分來描述：

- (a) 構造成分隨著常數平均值或常數趨勢；
- (b) 隨機、空間相關組成分；
- (c) 隨機、雜訊或剩餘機差項；

$$Z(X) = m(X) + \varepsilon'(X) + \varepsilon''$$

$m(X)$ ：主成分

$\varepsilon'$ ：猜測、局部改變、空間性質與  $m(X)$  剩餘項有關

$\varepsilon''$ ：剩餘項與空間性無關，平均為 0，變方為  $\delta^2$

Kriging 第一步是決定適合的函數 for  $m(X)$

$$E(Z(x) - Z(x+h)) = 0$$

$$E\{[Z(x) - Z(x+h)]^2\} = E\{[Z'(x) - Z'(x+h)]^2\} = 2r(h)$$

$$\hat{r}(h) = \frac{1}{2n} \sum_{i=1}^n [Z(x_i) - Z(x_i+h)]^2$$

$\hat{r}(h)$  對  $h$  作圖，稱為 *sample semi variogram*，

是一個重要的步驟，在內插時決定適當的權重。

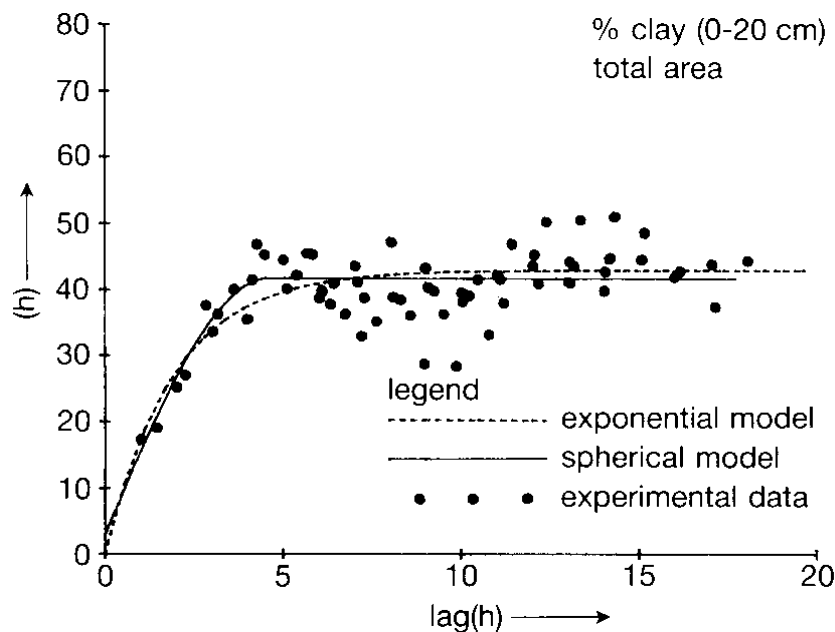


圖 8-11 顯示標準實驗土壤資料的 semivariogram

空間變化中方向效果：

semivariogram 將提供關於內插時應由那個樣品點 window 的形狀及方位問題的答案。

使用 semivariogram 內插

the fitted semivariogram 可用來決定局部內插所需的權重  $\lambda_i$

當  $\sum_{j=1}^n \lambda_j r(x_i x_j) + \varphi = r(x_i x_0)$  for all  $i$  時

$\hat{Z}(x_0)$  的最小變方 and is  $\sigma_e^2 = \sum_{j=1}^n \lambda_j r(x_j x_0) + \varphi$

Kriging 可找到較佳方法去估計內插權重，並提供關於機差的資訊。如果資料包括了長範圍趨勢，主成分  $m(x)$  (8-14 式) 應由 global polynomial trend surface 或用一組等級平均來模式化。

經常一組資料包括了局部趨勢，而與全部調查區不一致，這種局部趨勢反映了在地貌上不同的過程，在這些情形下，有些作者用 universal kriging 即 8-14 式中  $m(x)$  代表這些局部趨勢。因此剩餘式  $Z'(x) = Z(x) - m(x)$ ， $Z'(x)$  是區域化的變數。 $Z'(x)$  計算：一組已知資料，如果已知 semivariance，所以 drift 可求出 assumption；反之，亦然。

(a) drift 分析 (b) semivariogram 形式 (c) drift 操作時 sliding neighborhood

(a) drift：在一個特別方向，區域變數內系統性的增減。Drift 可出現在全尺度，在 universal kriging 重要，是連續的，改變緩慢，

可定義為  $m(x) = \sum_{\ell=1}^n a_{\ell} f_{\ell}(x)$

$a_{\ell}$ ：未知係數

$f_{\ell}(x)$ ：x 的整數權重

$n$ ：terms 數目

linear： $m(x) = a_0 + a_1 x$

quadratic： $m(x) = a_0 + a_1 x + a_2 x^2$

(b) semivariogram model 的選擇：

選用線性模型來模擬 residual 的 semivariogram 在接近起始處是線性的，以及在內插時只需 1-2 lag 距離資料。

Slope of semivariograms

$$b = \frac{k-1}{k-2} r_E(1)$$

$K$ ：相鄰樣品數

$r_E(1)$ ：在 lag 1 估計 semivariance

(c) the sliding neighborhood 之大小：

應用上推論一個區域包含 12-15 個樣品點是適當的。

Universal kriging 的應用及限制：

Universal kriging 可非常有用的應用於包含了定義良好局部趨勢的資料，地下水及平緩改變土地形態，卻可用這種方式成功的內插，但是土壤性質包含了大量局部變異的來源，就不容易。

區別真正的趨勢或採樣的變異：

*kriging* 複雜化：

semivariogram 的形式以 kriging 為中心，但並不容易確定 semivariogram 的估計，實際上在一個地區內是空間變異的真實估值。

總結：kriging 是一項進階的技術，很依賴統計定理及計算，特別是在製作地圖時，基本上，kriging 是一種理想的內插法，在估算過程中，敘述了空間影響的區域大小及形狀，不論是點或面，另一項好些是這個方法可估計在每個內插中伴隨而來之誤差，這是其他方法辦不到的。問題是出在移除變化的資料達到假設的情況，需要外加設備。

*Kriging* 擴張到大面積：

多數 kriging 用在小面積，頂多幾公頃而已，近來一些學者將其應用至大面積，因 kriging 是 scale-free 的技術，擴張至大面積並不會扭曲理論的假設，並使使用者了解這種變化與區域有關猶勝與局部範圍之關連。統計的不變性問題或一致性可與地景空間操作過程尺度及種類有關。

**Kriging 與其他內插技術的比較：**

一個好的 GIS 系統應包括一個範圍的內插技術，可允許使用者選擇最適當的方法。



表 8-3 內插方法之比較：提供使用者快速決定用那種方法。

方法	D：定 S：推	L：區域 G：域	A：突 G：漸進	外推	限制	利用於	輸出	計算	模式假設
眼	D	G	A	N	無重覆主觀	田野資料航照判釋	多邊形	無	直覺判斷區域內同質
邊緣法則	D	G	A	N	型態需定義並提供對人造地景優於天然地景	遙測影像	網格	中等	區域內同質
徐氏 (最近代理)	D	L	A	Y	誤差無法估計形狀依資料點分布而變	定性資料	多邊形	輕—中	最近區域為最佳資訊
趨勢面	S	G	G	N	邊緣效應大概輪廓多項誤差很少空間獨立性	廣面特性明顯其他方法的先前研究	網格點	輕—中	趨勢由多頁式解釋
傅立葉序列	S	G	G	N	不配適資料週期固定	週期性如山丘波浪人造物	網格點	中	對象具嚴厲的週期
曲線尺規	D	L	G	Y	誤差無法估計掩飾表面不確定性	平滑曲面	網格點	輕—中	完全平滑變化
移動平均	D	L	G	N	結果受資料點分布屬性、視窗大小影響變異假設誤差無法估計	快速產生適當等值線	網格點	中	連續等差表面
理想推估 kriging	S	L	G	Y	推估點性質之現實與理論問題計算資源大	詳細推估並要求誤差	網格點	大	本質假定(第一變異均質)區域均值更由連續面表示

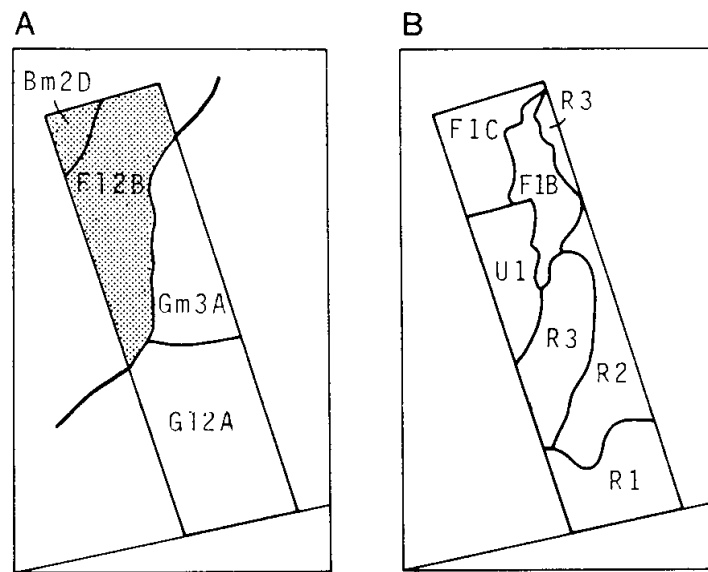


圖 8-12 Follonica 的土壤圖研究(a)1:50,000 尺度(b)1:25,000 尺度

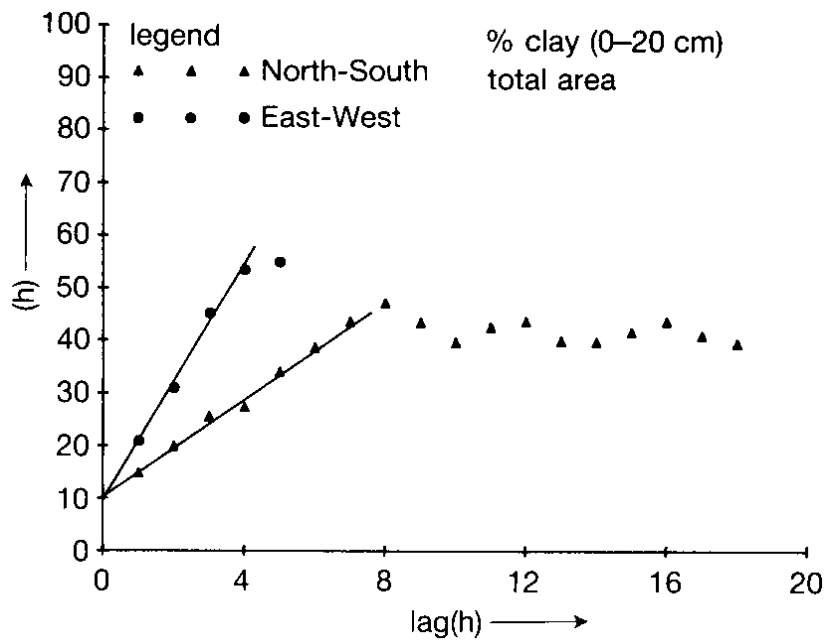


圖 8-13 Follonica 資料，以兩個方向來計算 clay (0~20cm)百分比的半變異圖

● 通過網格的短軸 ▲ 沿著網格長軸

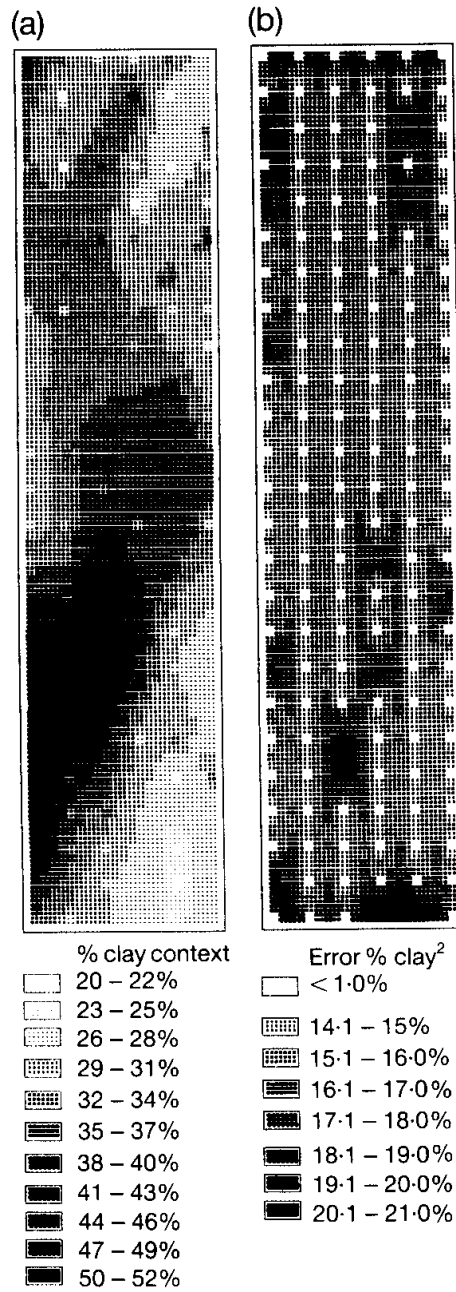


圖 8-14 資料為 Follonica 地區的高地土壤黏粒資料計算  
(a)以 37.5x37.5m 的網格點克利金法求得的圖(b)克利金變異圖

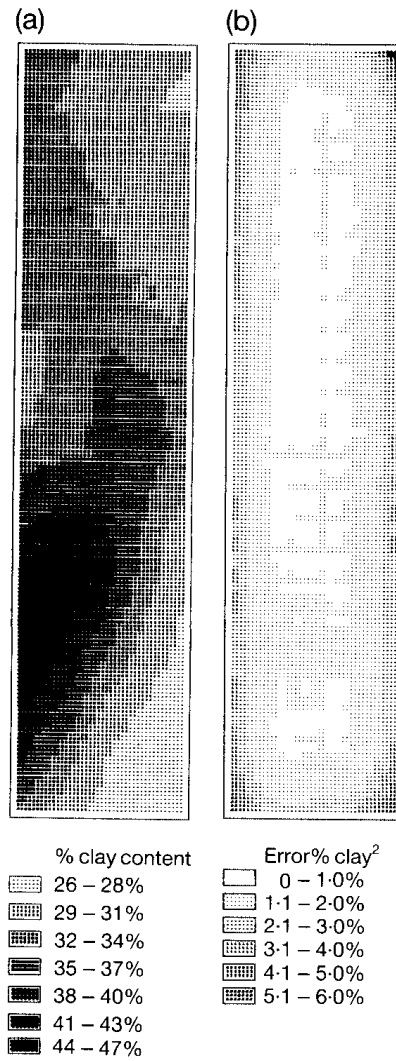


圖 8-15 計算資料來源(a)區塊克利金法 250×250m 的區塊，網格為 37.5×37.5m  
所得之圖(b)克利金變異圖

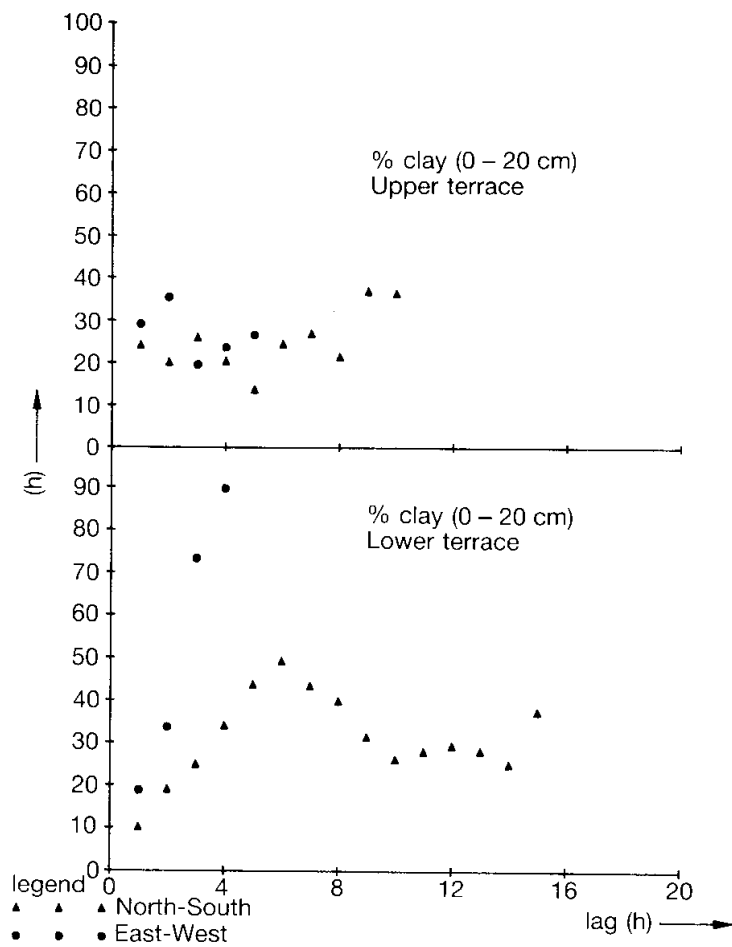


圖 8-16 Follonica 地高地黏粒百分比的半變異圖分為兩種主要單元  
• 平行的台地(高地)邊界(parallel with terrace boundary) ▲特殊的高地邊界

## 第九章 如何選擇 GIS

在前幾章介紹 GIS 的技術、結構與潛能，可以發現電腦整合技術和空間資料能有效地提供使用者做環境分析，尤其是電腦輔助製圖。

然而，在應用地理資訊系統前，必須先確定三件事：

- 1.你要建立或購買的系統，是否真正符合你技術上的需求；
- 2.因為 GIS 不便宜，而且需要專業人才，所以要有足夠的資金運用；
- 3.所採用 GIS 的種類，對於公司的工作影響甚大，並且在選擇系統前，應先盡可能將工作規劃完成。

**定義 GIS 的需求：**

1.使用的對象及使用者需求：我們可分成下列三級：

(a)級：使用者對工作可完全掌握：自動化的 GIS，對現有方法、效率的改變不大。

(b)級：使用者可掌握部分的工作：使用者對資訊的需求，只知道一部分。

(c)級：使用者完全不能掌握工作：每個工作都不同，使用者對資訊的需求，也不知道或是會改變的。

(a)級使用者是以製圖和資源調查部及實際使用資料的公司為代表，多在建立工作過程，他們有興趣的是電腦軟體公司。在歐洲他們佔有了市場的70%。

(b)級使用者包括了環境單位繪製地圖的部門，多數是系統調查的一個部門，他們的市場通常很小，許多電腦公司不會對這部分做重大投資，除非有特別發展時，才會作投資。

(c)級使用者是最困難的，包括了大學研究教學團體以及研究單位，他們用 GIS 當作工具，或作為發展新的 GIS 技術使用。

2.應用範圍：

對 GIS 系統種類的的需求，大多是以其工作的種類而決定，包括自然資源調查、資料利用、礦物搜尋及區別空間上與時間上尺度的需要。另一個重要的問題是 GIS 在安裝、設定上必須考慮應用電腦軟硬體的技术退化問題。電腦硬體的生命週期據估計約 5 年，軟體則較有彈性。每個 GIS 有其自己的軟體控制資料結構，如果軟體改變，資料結構必須重新建立。

3.技術選擇—向量或網格：

GIS 以 2 個方向發展：1.用向量模式(vector model)描述點、線、面；2.用網格模式(raster model)的 cell 矩陣及數學方程式來描述空間變異。

Vector 及 raster 方法可混合使用於資料庫中，這些方法可把 vector 轉換成 raster，反之亦然。把 raster 轉換成 vector 較為複雜，但圖形表現則較不令人滿意。在第 2 章已討論這二種方法的轉換，現在將其特點列在表 9-1 作比較：

表 9-1 GIS 中 vector 及 raster 資料結構之特長：

- 
- 1.使用 vector：以現象資料結構存檔。
  - 2.使用 vector：做路網分析。
  - 3.使用 vector：以高品質畫線的資料結構和展示。
  - 4.使用 raster：可製作快而便宜的地圖、容易進行疊圖及空間分析。
  - 5.需要跟地表面資料結合時，則以 raster 方法來模擬建立模式較容易。
  - 6.使用 raster 及 vector 結合：繪製高品質的線及面積塗色，畫線用 vector 格式，畫網格則用壓縮的 raster 結構，例如處理行長編碼或四分樹法。
  - 7.用壓縮的 vector 結構做數位地形模型較佳，且不會忽略高程的資料。
  - 8.raster-vector 及 vector-raster：把資料轉換成最適當的形式，俾供分析利用。
  - 9.展示可在 vector 及 raster 系統上操作，與儲存資料處理的資料結構無關。
- 

4.可運用財源：依可運用的資金可分成三級：

(a)財源充分—資金不受限制，則可購置理想的 GIS

(b)財源有限—要努力爭取經費以得到符合高品質需求

(c)財源受限—由於財政不夠，導致未來嚴重的問題

(a)級狀況是豪華的，可能只有在軍事上。多數希望能達(b)級；多數使用者，尤其是大學只達(c)級。

5.可利用的人員：

GIS 需要訓練過的人員來操作：

Low-skilled 職員：打字員、電腦操作員、數位化地圖人員，最後印圖之作圖者。

High-skilled 職員：分成 4 類：

管理者—每天使用 GIS、協調 GIS 與組織

聯絡人—需與使用者保持聯絡

技術人員—電腦製圖、程式員、系統建立者

科學家—環境科學家及應用 GIS 在研究上及應用問題尺度

現有的職員：(a)大量低技術(low-skilled)：沒有較高技術(high-skilled)

(b)大量低技術和一些較高技術的科學家及職員

(c)一些低技術和一些高技術的科學家及職員

(d)沒有或很少低技術：一些 high-skilled 科學家及職員

6.組織的情況：

要考慮其與組織其他部門之相互關係。

在組織內的 GIS 設備須配合不同來源的資料收集、處理方法，不同種類的產品。使用 GIS 設備可將一些傳統操作，如：彩色地圖的印製，取代不再使用人工操作。

GIS 比其取代的傳統方法，在短期內昂貴許多，但為了生存，公司必須使用新技術。而選擇提供使用電腦複雜的軟硬體，並不是件容易的事。

#### 7. GIS 組成的花費及人員需求：

建立 GIS 組成的實際成本隨著應用、大小及其來源而異（大型的繪圖機、數化板比小型的要昂貴許多），以 1985 年貨物價值計算：

1985 年	1998 年
低 < 10,000 美元	
中：10,000~50,000 美元	
高：50,000~250,000 美元	
極高 > 250,000 美元	

將 GIS 所需設備區分成輸入、輸出、分析等幾個部分對成本分析是有幫助的，這樣對各項操作所需人力、成本也再容易確認。表 9-2~9-6 列出其彼此關係，這些表整理出 GIS 所需之資料輸入、輸出、資料分析用的軟體、規劃與管理電腦等所需投資額及人事。

資料輸入：使用數化儀連接微電腦，離線建檔時，資料輸入並不昂貴。掃描儀在數化大量高品質線資料時是值得購買，非空間屬性資料用文字檔建檔最容易，連接空間資料至屬性表可用軟體進行投影轉換，在很便宜的在小電腦即可完成。資料輸入要正確以獲得不同來源，並確保資料庫架構及資料更新。

資料輸出：輸出有許多不同方式，輸出資料的品質要符合使用者需求。

資料分析所用之軟體：包括擷取、搜集、模式化、內插、統計分析、訓練職員等所使用的軟體或巨集。

規劃與管理：每個階層都要處理。

系統經理：每天規劃、做好資料提供者，並與軟體供應者保持接觸，確保系統運轉平順。

科學家：經營其規劃之工作。

聯絡員：明瞭 GIS 的角色，並需要確保財源、人力及給予精神上的支持。

主電腦：可以在每種電腦上建立其 GIS，明確選擇電腦的大小及類型。32 位元，小電腦 3-5Mbyte 內部記憶體。Disk memory 300-1,000Mbyte、1,600BPI tape drive 夠用了。由於技術進度目前電腦主機已進步到使用 32 位元處理器、64MB、硬碟容量亦不斷擴增。因為許多 GIS 操作不需大量計算，一個大而複雜的資料庫，硬碟儲存的速度及大小通常比實際過程的速度重要。硬碟存取的問題通常是一個重要因子，當 GIS 或 CAD/CAM 在大電腦使用時，分時(time-sharing)(讓系統設備上每個使用者均有獨自使用該電腦的感覺)嚴重限制了系統執行。



表 9-2 資料輸入 GIS 所需的投資額和人事

操作 GIS	成本	人事			
		低技術	高技術		
			技術的	科學的	管理的
微電腦離線數化	低/中	+	—	—	(+)
變換地理系統線上數化	高	+	+	—	(+)
離線掃描(高品質)+向量轉換	非常高	(+)	*	—	(+)
ASCII 檔屬性輸入	低	+	—	—	(+)
建立資料庫內位相網路	低/中	(+)	+	—	(+)
連結地理和非地理資料	低	(+)	+	—	(+)
置入衛星影像	低	+	+	—	(+)
衛星影像之幾何轉換	低()高()	+	+	—	(+)
資料庫和文件紀錄稽查	低/中	+	+	(+)	+

表 9-2~9-6 符號說明：+必需的 \*絕對需要的 -不需要的 (+)想要的但非絕對需要的

表 9-3 資料從 GIS 輸出所需的投資額和人事

硬體操作	成本	人事			
		低技術	高技術		
			技術的	科學的	管理的
低品質矩陣繪圖儀	低	+	—	—	—
筆式繪圖儀(簡單型)	低	+	—	—	—
噴墨繪圖儀(低品質)	低	+	—	—	—
噴墨繪圖儀(高品質)	中/高	(+)	+	—	(+)
鐳射繪影單行	高	—	*	—	(+)
相 繪圖儀/抄寫	高/非常高	—	*	+	(+)
矩陣繪圖儀(品質好的)	中/高	+	+	—	(+)
矩陣繪圖儀(品質極好的)	高	+	+		(+)
磁帶機	低	+	—	—	(+)
以 hard-copy 設備	中/高	+	+	(+)	(+)
地理色彩銀幕上製圖繪製和標準	中/高	—	+	*	+
高精密度影件記錄(全彩)	高	—	+		(+)

符號：和表 9-2 定義相同

表 9-4 資料分析使用軟體所需的投資額和人事

操作 GIS	成本	人事			
		低技術	高技術		
			技術的	科學的	管理的
簡單資料擷取	低	+	—	—	(+)
使用相關資料庫管理系統	高	(+)	*	+	(+)
地理系統相互作用	高/非常高	(+)	+	—	(+)
內插和畫等值線	中/高	—	+	+	(+)
地圖套疊	中/高	(+)	+	+	(+)
統計分析	低/中	—	(+)	*	(+)
網格影像處理分析	中/高/非常高	—	+	*	(+)
數值地形模型	中/高	—	+	*	(+)
特殊軟體	依型面別	—	+	*	(+)
軟體庫	低/中/高	—	+	*	(+)

符號：和表 9-2 定義相同

表 9-5 對 GIS 規劃和管理所需的投資額和人事

操作 GIS	人事				
	低技術	高技術			
		技術的	科學的	管理的	對外聯繫
組織內部規劃工作流程	(+)	+	+	*	—
與資料提供者聯繫	—	+	+	*	*
與使用者聯繫	—	+	+	*	*
與硬體和軟體提供者聯繫	—	*	*	*	+
與其他製圖機構聯繫	—	+	*	*	+

符號：和表 9-2 定義相同

表 9-6 GIS 電腦主機所需的投資額和人事

電腦類別	人事				
	成本	高技術			
		低技術	技術的	科學的	管理的
微電腦（個人）	低	+	(+)	(+)	(+)
微電腦（16 位元硬碟）	中	+	+	(+)	(+)
小型電腦	高/非常高	+	+	+	(+)
工作站	通常沒有應用於 GIS				

符號：和表 9-2 定義相同

**在設定 GIS 時應遵循的步驟：**

- 1.將工作情形建立一份詳細清單：了解未來需求，不只滿足經費需要，更發展未來成長所需設備，細節包括每個工作量，資料庫大小，輸出的正確性、品質及需要的人力。
- 2.鑑定滿足設備所需的軟、硬體種類。
- 3.尋找所需軟體的供應商，如：附錄 2 所列。注意一些特別便宜，未獲支援的套裝軟體(package)，它雖然看來吸引人，但可能使你的公司花費許多人/月去學習操作，長期下來，花費將超過一個昂貴而品質好的產品。
- 4.考慮電腦所需的硬碟，解答軟硬體的問題可用看守系統(turn-key system)：a turn key system 把軟硬體結合在一起，以解決特殊需求，如遙測影像分析做電腦輔助設計(CAD)和製圖。
- 5.準備文件、詳列需求，使供應商提出規劃建議書。建議書應符合需求，並且需經過「基準點測試(benchmark test)」的處理。  
benchmark test：是一種工作模組，公司希望要做的工作先委託潛在的供應商，廠商被要求完成這項測試，所需時間(如：數化、多邊形套疊或投影轉變)、使用容易度、位置正確性，這些 test 可以詳細、也可以濃縮。
- 6.依價格與績效選擇一個系統：員工訓練計畫，通常在 GIS 完全操作前要 9-12 個月要建立，甚至更長，可用一些小的試驗計畫來訓練員工，以展示系統功能。